frontiers | Research Topics

Reviews in recommender systems 2022

Edited by
Dominik Kowald, Deqing Yang and
Emanuel Lacić

Published in
Frontiers in Big Data

FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-4766-3
DOI 10.3389/978-2-8325-4766-3

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Reviews in recommender systems: 2022

Topic editors

Dominik Kowald — Know Center, Austria

Deqing Yang — Fudan University, China

Emanuel Lacić — Infobip d.o.o., Croatia

Citation

Kowald, D., Yang, D., Lacić, E., eds. (2024). *Reviews in recommender systems: 2022*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4766-3

Table of contents

frontiers | Frontiers in Big Data

TYPE Editorial
PUBLISHED 02 April 2024
DOI 10.3389/fdata.2024.1384460

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Huan Liu,
Arizona State University, United States

*CORRESPONDENCE
Dominik Kowald
✉ dkowald@know-center.at

RECEIVED 09 February 2024
ACCEPTED 19 March 2024
PUBLISHED 02 April 2024

CITATION
Kowald D, Yang D and Lacic E (2024) Editorial: Reviews in recommender systems: 2022. *Front. Big Data* 7:1384460.
doi: 10.3389/fdata.2024.1384460

COPYRIGHT
© 2024 Kowald, Yang and Lacic. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Reviews in recommender systems: 2022

Dominik Kowald[1,2]*, Deqing Yang[3] and Emanuel Lacic[4]

[1]Know-Center GmbH, Graz, Austria, [2]Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, Austria, [3]School of Data Science, Fudan University, Shanghai, China, [4]Infobip, Zagreb, Croatia

KEYWORDS

recommender systems, review, fairness, privacy, collaborative filtering

Editorial on the Research Topic
Reviews in recommender systems: 2022

1 Introduction

Nowadays, recommender systems are one of the most widely used instantiations of machine learning and artificial intelligence. Thus, these systems accompany us in our daily online experience and have become an integral part of our digital life for supporting us in finding relevant information in information spaces that are too big or complex for manual filtering (Ricci et al., 2010; Burke et al., 2011; Jannach et al., 2016). Since the first deployments of recommendation algorithms (Resnick et al., 1994; Resnick and Varian, 1997), recommender systems analyze past usage behavior (e.g., clicks or ratings) in order to build user models, and to suggest items to users. Recommender systems are employed in various domains, ranging from entertainment domains, such as music (Lex et al., 2020; Schedl et al., 2021) and movies (Harper and Konstan, 2015), to more critical domains such as the job market (Lacic et al., 2020). Apart from that, different types of algorithms have been employed to develop recommender systems, ranging from collaborative filtering (Ekstrand et al., 2011), content-based filtering (Lops et al., 2010), hybrid approaches (Burke, 2002), theory-driven algorithms [e.g., based on cognitive models (Lacic et al., 2014; Kowald et al., 2015)], to neural approaches (Zhang et al., 2019; Chen et al., 2023).

The aim of the "*Reviews in recommender systems*" Research Topic is to highlight recent advances in the broad field of recommender systems, including important topics such as fairness (Kowald et al., 2020; Wang et al., 2023), privacy (Friedman et al., 2015; Muellner et al., 2021), and multi-stakeholder objectives (Abdollahpouri and Burke, 2019), while emphasizing novel directions and possibilities for future research. In total, this Research Topic consists of nine review articles surveying the literature in a specific subfield of recommender systems. More concretely, the editors of this Research Topic have been able to accept six full-length and three mini review articles. The following section gives a short overview of these articles.

2 Research Topic content

In a mini review article, Müllner et al. surveyed the current landscape of differential privacy in collaborative filtering-based recommender systems. In total, the authors have reviewed 26 publications, and found that in most cases, differential privacy is applied to the user representation (i.e., the input data of the recommender system) rather than to recommendation model updates or to phases after the training. Additionally, the authors stated that most papers investigate differential privacy on datasets gathered from MovieLens and Last.fm, and thus, that more research is needed for privacy-aware recommender systems in sensitive domains such as the job market or finance. Next, Jannach and Abdollahpouri explore the multifaceted landscape of multi-objective recommender systems, identifying the need to balance diverse and often conflicting objectives such as user satisfaction, stakeholder interests, and long-term goals of stakeholders. The authors present a taxonomy categorizing these objectives into recommendation quality, multi-stakeholder perspectives, temporal considerations, user experience, and system engineering challenges. The study illustrates the complexity of optimizing recommender systems in real-world applications, emphasizing the importance of addressing multiple objectives to enhance recommendation relevance, diversity, and overall system effectiveness.

Banerjee et al. delve into the challenges and potential strategies for ensuring fairness in Tourism Recommender Systems (TRS), emphasizing the multi-stakeholder nature of these systems. They categorize stakeholders based on fairness criteria, review state-of-the-art research from various perspectives, and highlight the complexities of balancing individual and collective interests. The paper concludes that achieving fairness in TRS involves navigating trade-offs between stakeholder interests, illustrating the necessity for innovative solutions that consider the environmental impact and societal concerns alongside traditional user and provider objectives. In the next mini-review, Loepp investigates the increasingly prevalent multi-list user interfaces in recommender systems, particularly focusing on carousel-based interfaces like those used by Netflix and Spotify. The review highlights the scarcity of research on optimizing these carousels for user interaction and satisfaction, despite their common use. Based on 18 reviewed research papers, the author identifies gaps in understanding user behavior and interface design, and proposes future research directions to enhance user experience through improved design and personalization of carousel recommendations.

Kumar et al. provide an in-depth review of fairness in recruitment-related recommender systems (RRSs), dissecting the balance between technical advancements and legal compliance. They delve into various fairness definitions (e.g., demographic parity), metrics (e.g., false positive rates between different demographic groups), and debiasing strategies (e.g., post-processing to alter the algorithm's output to ensure fairness) as well as compare them to existing EU and US employment laws. The survey spotlights the nuanced challenges of mitigating algorithmic bias and discrimination within RRSs, advocating for a multidisciplinary approach to develop more equitable and legally compliant hiring technologies. Additionally, Felfernig et al. explore the potential of recommender systems to support the achievement of the 17 United Nations' Sustainability Development Goals (SDGs). The review addresses the utilization of AI to recommend actions and alternatives aligned with sustainability objectives. The paper discusses various recommender system types, their application across all SDGs, as well as identifies open research issues for future exploration. The authors show the significance of recommender systems in promoting sustainability, offering both current insights and directions for ongoing research.

In this mini-review, Duricic et al. explore the integration of beyond-accuracy metrics (i.e., diversity, serendipity, and fairness) into recommender systems based on Graph Neural Networks (GNNs). They emphasize the importance of these metrics in enhancing user satisfaction, beyond mere accuracy. Furthermore, they examine recent advancements and methodologies in GNNs that address these dimensions, highlighting the balance between recommendation accuracy and beyond-accuracy objectives. Next, Lubos et al. present a review of state-of-the-art video recommender systems (VRS), covering a broad range of algorithms, applications, and unresolved research challenges in the field. They delve into various approaches to VRS, including content-based, collaborative filtering, and hybrid systems, and discuss the importance of diverse content representations and evaluation metrics. Based on the analysis of 6 different application domains, they highlight the potential for future advancements in VRS, emphasizing the need for innovative solutions to improve the accuracy and effectiveness of personalized video recommendations, thereby serving as a valuable resource for both researchers and practitioners in the video domain. Finally, Uta et al. offer a comprehensive overview of knowledge-based recommender systems, distinguishing them from traditional collaborative and content-based approaches by their ability to utilize semantic user preferences, item knowledge, and recommendation logic. These systems are particularly beneficial for complex item types, as they can dynamically adapt to user preferences through dialogue and constraint-based recommendations. The review also identifies future research directions, emphasizing the integration of knowledge-based technologies in recommender systems.

Taken together, across all review articles, we see that beyond-accuracy objectives and trustworthiness aspects of recommender systems are currently of high interest in the recommender systems research community. This includes aspects related to fairness, bias, privacy, diversity, serendipity, sustainability, multi-stakeholder objectives, and user interface choices. We hope that the review articles presented in this Research Topic will inform future research endeavors in this field.

Author contributions

DK: Writing – original draft, Writing – review & editing. DY: Writing – original draft, Writing – review & editing. EL: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the FFG COMET funding program.

Conflict of interest

DK was employed by Know-Center GmbH. EL was employed by Infobip.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdollahpouri, H., and Burke, R. (2019). Multi-stakeholder recommendation and its connection to multi-sided fairness. *arXiv* [Preprint]. arXiv:1907.13158. doi: 10.48550/arXiv.1907.13158

Burke, R. (2002). Hybrid recommender systems: survey and experiments. *User Model. User Adapt. Interact*. 12, 331–370. doi: 10.1023/A:1021240730564

Burke, R., Felfernig, A., and Göker, M. H. (2011). Recommender systems: an overview. *AI Mag*. 32, 13–18. doi: 10.1609/aimag.v32i3.2361

Chen, X., Yao, L., McAuley, J., Zhou, G., and Wang, X. (2023). Deep reinforcement learning in recommender systems: a survey and new perspectives. *Knowl. Based Syst*. 264:110335. doi: 10.1016/j.knosys.2023.110335

Ekstrand, M. D., Riedl, J. T., and Konstan, J. A. (2011). Collaborative filtering recommender systems. *Found. Trends Hum.-Comput. Interact*. 4, 81–173. doi: 10.1561/1100000009

Friedman, A., Knijnenburg, B. P., Vanhecke, K., Martens, L., and Berkovsky, S. (2015). "Privacy aspects of recommender systems," in *Recommender Systems Handbook*, eds F. Ricci, L. Rokach, and B. Shapira (Cham: Springer), 649–688. doi: 10.1007/978-1-4899-7637-6_19

Harper, F. M., and Konstan, J. A. (2015). The movielens datasets: history and context. *ACM Trans. Interact. Intell. Syst*. 5, 1–19. doi: 10.1145/2827872

Jannach, D., Resnick, P., Tuzhilin, A., and Zanker, M. (2016). Recommender systems—beyond matrix completion. *Commun. ACM* 59, 94–102. doi: 10.1145/2891406

Kowald, D., Schedl, M., and Lex, E. (2020). "The unfairness of popularity bias in music recommendation: a reproducibility study," in *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II 42* (Cham: Springer), 35–42. doi: 10.1007/978-3-030-45442-5_5

Kowald, D., Seitlinger, P., Kopeinik, S., Ley, T., and Trattner, C. (2015). "Forgetting the words but remembering the meaning: modeling forgetting in a verbal and semantic tag recommender," in *Mining, Modeling, and Recommending 'Things' in Social Media* (Cham: Springer), 75–95. doi: 10.1007/978-3-319-14723-9_5

Lacic, E., Kowald, D., Seitlinger, P. C., Trattner, C., and Parra, D. (2014). "Recommending items in social tagging systems using tag and time information," in *Proceedings of the 1st Social Personalization Workshop co-located with the 25th ACM Conference on Hypertext and Social Media* (New York, NY: Association of Computing Machinery), 4–9.

Lacic, E., Reiter-Haas, M., Kowald, D., Reddy Dareddy, M., Cho, J., Lex, E., et al. (2020). Using autoencoders for session-based job recommendations. *User Model. User Adapt. Interact*. 30, 617–658. doi: 10.1007/s11257-020-09269-1

Lex, E., Kowald, D., and Schedl, M. (2020). Modeling popularity and temporal drift of music genre preferences. *Trans. Int. Soc. Music Inf. Retr*. 3, 17–30. doi: 10.5334/tismir.39

Lops, P., De Gemmis, M., and Semeraro, G. (2010). Content-based recommender systems: state of the art and trends. *Recommender Systems Handbook*, eds F. Ricci, L. Rokach, and B. Shapira (Cham: Springer), 73–105. doi: 10.1007/978-0-387-85820-3_3

Muellner, P., Kowald, D., and Lex, E. (2021). "Robustness of meta matrix factorization against strict privacy constraints," in *Advances in Information Retrieval: 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28–April 1, 2021, Proceedings, Part II 43* (Cham: Springer), 107–119. doi: 10.1007/978-3-030-72240-1_8

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). "Grouplens: an open architecture for collaborative filtering of netnews," in *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work* (New York, NY: ACM), 175–186. doi: 10.1145/192844.192905

Resnick, P., and Varian, H. R. (1997). Recommender systems. *Commun. ACM* 40, 56–58. doi: 10.1145/245108.245121

Ricci, F., Rokach, L., and Shapira, B. (2010). "Introduction to recommender systems handbook," in *Recommender Systems Handbook*, eds F. Ricci, L. Rokach, and B. Shapira (Cham: Springer), 1–35. doi: 10.1007/978-0-387-85820-3_1

Schedl, M., Bauer, C., Reisinger, W., Kowald, D., and Lex, E. (2021). Listener modeling and context-aware music recommendation based on country archetypes. *Front. Artif. Intell.* 3:508725. doi: 10.3389/frai.2020.508725

Wang, Y., Ma, W., Zhang, M., Liu, Y., and Ma, S. (2023). A survey on the fairness of recommender systems. *ACM Trans. Inf. Syst*. 41, 1–43. doi: 10.1145/3547333

Zhang, S., Yao, L., Sun, A., and Tay, Y. (2019). Deep learning based recommender system: a survey and new perspectives. *ACM Comput. Surv*. 52, 1–38. doi: 10.1145/3158369

TYPE Original Research
PUBLISHED 22 March 2023
DOI 10.3389/fdata.2023.1157899

Check for updates

OPEN ACCESS

EDITED BY
Dominik Kowald,
Know Center, Austria

REVIEWED BY
Peter Müllner,
Know Center, Austria
Markus Reiter-Haas,
Graz University of Technology, Austria

*CORRESPONDENCE
Dietmar Jannach
✉ dietmar.jannach@aau.at

SPECIALTY SECTION
This article was submitted to Recommender Systems, a section of the journal Frontiers in Big Data

RECEIVED 03 February 2023
ACCEPTED 07 March 2023
PUBLISHED 22 March 2023

CITATION
Jannach D and Abdollahpouri H (2023) A survey on multi-objective recommender systems. *Front. Big Data* 6:1157899. doi: 10.3389/fdata.2023.1157899

COPYRIGHT
© 2023 Jannach and Abdollahpouri. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A survey on multi-objective recommender systems

Dietmar Jannach[1]* and Himan Abdollahpouri[2]

[1]Department of Artificial Intelligence and Cybersecurity, University of Klagenfurt, Klagenfurt, Austria,
[2]Spotify, Inc., New York, NY, United States

Recommender systems can be characterized as software solutions that provide users with convenient access to relevant content. Traditionally, recommender systems research predominantly focuses on developing machine learning algorithms that aim to predict which content is relevant for individual users. In real-world applications, however, optimizing the accuracy of such relevance predictions as a single objective in many cases is not sufficient. Instead, multiple and often competing objectives, e.g., long-term vs. short-term goals, have to be considered, leading to a need for more research in multi-objective recommender systems. We can differentiate between several types of such competing goals, including *(i)* competing recommendation quality objectives at the individual and aggregate level, *(ii)* competing objectives of different involved stakeholders, *(iii)* long-term vs. short-term objectives, *(iv)* objectives at the user interface level, and *(v)* engineering related objectives. In this paper, we review these types of multi-objective recommendation settings and outline open challenges in this area.[1]

KEYWORDS

recommender systems, evaluation, multistakeholder recommendation, beyond-accuracy optimization, short-term and long-term objectives

1. Introduction

Generically defined, recommender systems can be characterized as *software solutions that provide users convenient access to relevant content*. The types of conveniences that such systems provide can be manifold. Historically, recommender systems were mainly designed as information filtering tools, like the early GroupLens system (Resnick et al., 1994) from 1994. Later on, various other ways were investigated how such systems can create value, e.g., by helping users *discover* relevant content, by providing easy access to related content (e.g., accessories), or by even taking automatic action like creating and starting a music playlist.

While recommender systems can serve various purposes and create value in different ways (Jannach and Zanker, 2021), the predominant (implicit) objective of recommender systems in literature today can be described as "guiding users to relevant items in situations where there is information overload," or simply "finding good items" (Herlocker et al., 2000; Manouselis and Costopoulou, 2007; Cacheda et al., 2011; Kamishima et al., 2018). The most common way of operationalizing this information filtering problem is to frame the recommendation task as a supervised machine learning problem. The core of this problem is to learn a function from noisy data, which accurately predicts the *relevance* of a given item for individual users, sometimes also taking contextual factors into account.

Although the actual relevance of recommended items can be assessed in different ways (Gunawardana and Shani, 2015), data-based offline experiments dominate the research landscape. In the early years, rating prediction was considered a central task of a recommender, and the corresponding objective was to minimize the mean absolute error (MAE), see Shardanand and Maes (1995) for work using MAE in 1996. Nowadays, item

1 This paper is an extension of our previous work presented in Jannach (2022).

ranking is mostly considered to be more important than rating prediction, and a variety of corresponding ranking accuracy measures are used today.

While the metrics changed over time, the research community has been working on optimizing relevance predictions in increasingly sophisticated ways for almost 30 years now. The main objective of such research is to minimize the relevance prediction error or to maximize the accuracy of the recommendations. The underlying assumption of these research approaches is that better relevance predictions lead to systems that are more valuable for their users. This seems intuitive for many practical applications because a better algorithm should surface more relevant items in the top-N lists shown to users.

Such an assumption might however not always be true, and it was pointed out many years ago that "being accurate is not enough" (McNee et al., 2006) for a recommender system to be successful. A recommender system might for example present users with obvious recommendations, e.g., recommending new Star Wars sequels to a Star Wars lover. The prediction error for such recommendations might be even close to zero. But so will the value of the recommendations to users, who most probably know these movies already. Observations like this led to a multitude of research efforts on "beyond-accuracy" measures like diversity, novelty, or serendipity, see Bradley and Smyth (2001) for an early work from 2001.

Such beyond-accuracy measures typically compete with accuracy measures (Shi, 2013; Isufi et al., 2021), leading to the problem that multiple objectives have to be balanced when serving recommendations. Which beyond-accuracy dimensions are relevant for a given setting and how much weight should be given to the competing objectives in practice depends on application-specific aspects and in particular on the purpose the recommender is intended to serve (Jannach and Adomavicius, 2016).

Historically, when considering the purpose of a recommender system, the focus of the research was on the value of such a system for *consumers*. Only in recent years, more attention has been paid to the fact that recommender systems in practice factually serve some business or organizational objectives. Considering these platform and item provider-side aspects, therefore, requires that we see recommendation as a problem where the interests and objectives of multiple stakeholders must be considered (Abdollahpouri et al., 2020; Abdollahpouri and Burke, 2022), often also taking different optimization time horizons into account. In Abdollahpouri et al. (2020), the authors emphasize different types of stakeholders in a recommendation environment, namely, consumers, providers, and the recommendation platform. Plus, there can also be side stakeholders such as society. An ideal recommender system operating in a multi-stakeholder environment should aim to balance the objectives of different stakeholders to ensure all stakeholders are satisfied to a certain extent.

Overall, while being able to predict the relevance of individual items for users remains to be a central and relevant problem, considering only one type of objective, i.e., prediction accuracy, and the corresponding metrics may be too simplistic and ultimately limit the impact of academic research efforts in practice. Unfortunately, while we observed an increased research interest in beyond-accuracy metrics during the last 10 years, a large fraction of published works today focuses exclusively on accuracy or a rather limited set of other quality-related metrics. Therefore, one important way to escape the limitations of current research practice is to consider multiple types of optimization goals, stakeholder objectives and their trade-offs in parallel (Jannach and Bauer, 2020). Next, in Section 2, we will discuss various forms of multi-objective recommender systems found in the literature. To the best of our knowledge, the taxonomy we provide in this paper is the first in giving a holistic view of the landscape of multi-objective recommender systems. A recent survey on the topic by Zheng and Wang (2022) focuses largely on the specifics of existing technical approaches to balance multiple optimization objectives and discuss which approach is suitable for which class of problems. We refer readers to this valuable survey on technical aspects. Our present work in contrast aims to provide a more holistic picture of the various forms of multi-objective recommendation problems.

2. A taxonomy of multi-objective recommendation settings

In this section, we will first provide a high-level overview of a taxonomy of multi-objective recommendation settings and then discuss the individual components and representative examples in more depth.

2.1. Definition and taxonomy overview

On a very general level, we can define that "*a multi-objective recommender system (MORS) as a system designed to jointly optimize or balance more than one optimization goal.*" Figure 1 provides a taxonomy of different types of multi-objective recommendation settings.

We differentiate between five main *types* of objectives:

- **Quality objectives:** Various aspects that can contribute to the quality of recommendations presented to users, including relevance (accuracy), diversity, or novelty. In many cases, these quality objectives are assumed to be competing.
- **Multistakeholder objectives:** Recommender systems are usually designed with the goal of creating value both for consumers, service providers (also called *recommendation platforms*), and maybe other stakeholders such as item suppliers. Challenges for example arise when the best (most relevant) recommendations for the consumer are not the most valuable ones from the perspective of other involved stakeholders.
- **Time horizon objectives:** Recommendations can both impact the short-term and the long-term behavior of users. In the short term, recommendations are designed to help users to find relevant content and/or to influence their choices. Recommendations can however also have longitudinal effects, both positive ones (such as trust building toward the platform) or negative ones (such as filter bubbles) (Pariser, 2011), and again, long-term and short-term objectives may be competing.
- **User experience objectives:** There are various design options and potential trade-offs when developing the user experience

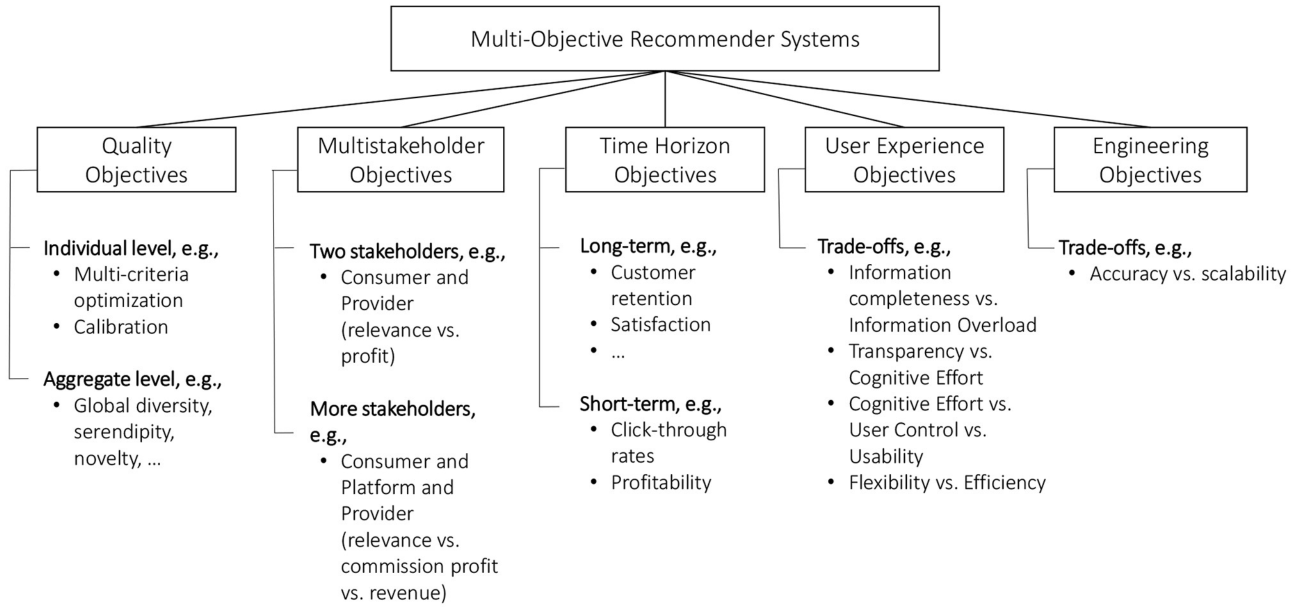

FIGURE 1
Taxonomy of different types of multi-objective recommendation settings.

of a recommender system. For example, one might try to reduce the cognitive load for users by limiting the amount of information that is presented, e.g., in terms of the number of choices. On the other hand, some users, sometimes referred to as "maximizers" (Schwartz et al., 2002) may instead prefer to see the full spectrum of options before making a decision.

- **Engineering objectives:** Finally, there may be trade-offs regarding engineering (or: system) related aspects. Modern machine learning models can for example be costly to train and challenging to debug. In such situations, it has to assessed if the investments in more complex solutions pay off in practice.[2]

We emphasize that the objectives in the described categories are not mutually exclusive, and in many cases, there are dependencies between the objectives in practice. This may not be immediately apparent from the academic literature, which historically largely focuses on quality objectives. In practical settings, however, the impact of the recommendations on the relevant Key Performance Indicators (KPIs) of the recommendation service provider will almost always be part of the optimization objectives as well.

Moreover, as mentioned, in many cases, the objectives both within a category and across categories can be competing and represent a trade-off. Dealing with such trade-offs is a common target in academic literature, in which most evaluations are done offline, i.e., based on historical data and without users in the loop. In such settings, the goal is then to find a balance between two or more computational metrics, e.g., diversity and accuracy. Limited research unfortunately exists that examines potential trade-offs through real-world experiments. The simulation study in Mehrotra et al. (2018) is an example of a work that is based on real-world A/B test log data, which indicates that increasing the system's fairness may lead to higher user satisfaction and engagement in practice. Also, when considering short-term and long-term objectives, taking measures to increase interactivity and engagement with the system in the short term is sometimes considered beneficial for customer retention in the long run (Gomez-Uribe and Hunt, 2015).

We discuss the different elements of our taxonomy and selected representative works next.

2.2. Recommendation quality objectives

Under this category, we subsume problem settings where more than one quality objective of recommendations for users must be considered. We can differentiate between the system considering such objectives at the level of individual users or at an aggregate level, i.e., for the entire user base.

2.2.1. Individual level

At the individual level, consumers can have specific (short-term) preferences, e.g., regarding item features that should be considered in parallel. For instance, a user of a hotel booking platform might be interested in a relatively cheap hotel, which in addition is in close proximity to the city center. In such a situation, the user has multiple criteria in mind for picking the ideal item and the goal of the recommender system is to balance these criteria and recommend items to the user that match the desired criteria as much as possible.

2 A prime example in this context is that Netflix never put the winning solution of their Netflix Prize Challenge into production, see https://www.wired.com/2012/04/netflix-prize-costs/.

A central problem for a recommender system in such situations is to acquire or derive the user's preferences for the different dimensions. In many cases, and in various early systems like the 1997 "FindeMe" approach to assisted browsing (Burke et al., 1997), preference elicitation is done in an *interactive* or *conversational* approach, see Gao et al. (2021); Jannach et al. (2021) for recent surveys on the topic. The acquisition of the user preferences can be done in different ways, e.g., through pre-defined dialog paths (e.g., Jannach, 2004), through statically or dynamically proposed *critiques* on item features (e.g., Chen and Pu, 2012), or, as done in most recent works, through natural language interactions (e.g., Li et al., 2018). A variety of alternative approaches were proposed as well, e.g., based on the Analytic Hierarchy Process (AHP), e.g., Liu and Shih (2005).[3] On a general level, such interactive recommendation systems, therefore, support their users in a Multi-Criteria Decision-Making (MCDM) process (Triantaphyllou, 2000; Manouselis and Costopoulou, 2007).

Various technical approaches can be used to derive a set of suitable recommendations once the preferences are acquired. In *constraint-based* systems, for example, explicitly specified rules are commonly used which filter out items that do not match the user preferences. In *case-based systems*, similarity functions play a central role in item retrieval. And in natural-language based systems sentiment analysis can for example be used to derive the user's preferences toward certain items or item features, and these preferences may then be fed into a collaborative filtering algorithm (Smyth, 2007; Felfernig et al., 2015; Li et al., 2018). In particular, in the case of constraint-based systems, the situation may occur that none of the items in the catalog fulfills all specified preferences. For example, assume a user is only interested in hotel rooms cheaper than $100 per night and in less than 5 kilometers from the city center. If no hotel room matches such constraints, the algorithm can relax some of the constraints so a set of recommendations that partially matches the user's criteria can be returned (Felfernig et al., 2015). Furthermore, methods like Multi-Attribute Utility Theory can be applied to rank the remaining candidates (Huang, 2011).

Besides approaches that interactively acquire the user preferences regarding certain item features, another line of research exists that is based on collaborative filtering and on *multi-criteria item ratings*. In such approaches (Adomavicius and Kwon, 2015), users are not expected to specify their preferences for different item features in general but are assumed to rate features of specific items. For example, in the tourism recommendation domain, they might assess a given hotel in dimensions such as value for money, cleanliness, or friendliness of the staff. This more fine-grained preference information can then be used in specifically-extended collaborative filtering approaches, e.g., Adomavicius and Kwon (2007) and Jannach et al. (2012).

A different way to take into account the often multi-faceted nature of individual user preferences is called *calibration*. In these approaches, the idea is not to find items that match user preferences in certain item-specific dimensions but to match past user preferences with respect to certain meta-level properties of the recommendation lists such as diversity. For instance, if for a user of a video streaming platform, interest in various genres was observed, a calibrated recommender system may try to generate a set of item suggestions that reflects this diversity of the user interests.

In an early work, Oh et al. (2011) tried to align the recommendations with the past popularity tendencies of a user where the authors tried to rerank the recommendation lists such that the distribution of the popularity of items in the recommended list to each user, matches their historical tendency toward such items. Later, Jugovac et al. (2017) extended the approach for multiple optimization objectives where authors tried to jointly optimize the relevance of the recommended items along with some additional quality factors such as list diversity, item popularity, and item release years. A more formal characterization of the calibration was introduced by Steck in Steck (2018) who proposed an approach for reranking the recommendations such that the final list is both relevant and also matched the genre preference of the users. Similarly, Abdollahpouri et al. (2021) represents another recent work in that direction where authors aim to tackle the popularity bias problem in recommender systems by reranking the recommendation lists generated for each user such that it has both high relevance and is also in line with the historical popularity tendency of the users.[4] Overall, in most cases, the central idea of calibration approaches is to match two distributions of some aspect of the recommended items. An alternative optimization goal was used in Jannach et al. (2015a) for the music domain, where the objective was to find musically coherent playlist continuations while preserving prediction accuracy.

2.2.2. Aggregate level

The majority of published research on balancing different recommendation quality aspects targets the aggregate level. The objective of such works is to balance the recommendations for the entire user base, the corresponding metrics are therefore usually averages.[5] The most common beyond-accuracy measures in the literature include diversity, novelty, serendipity, catalog coverage, popularity bias, or fairness, see, e.g., Adomavicius and Kwon (2012), Kaminskas and Bridge (2016), Vargas and Castells (2011), Abdollahpouri et al. (2017), and Ekstrand et al. (2022). Most commonly, the goal is to balance accuracy with exactly one of these measures, assuming that there is a trade-off between these quality factors. Increasing diversity is for example commonly assumed to have a negative impact on accuracy metrics. A few works exist which consider more than two factors. In an earlier work in this area (Rodriguez et al., 2012), the authors describe an effort to build a *talent recommendation* system at LinkedIn, which not only considers the semantic match between a candidate profile and a

3 See also He et al. (2016) and Jugovac and Jannach (2017) for surveys on interactive recommender systems.

4 While such a calibration approach turns out to be effective to mitigate popularity bias on the individual level, it may be limited in terms of reducing this bias across an entire user population (Klimashevskaia et al., 2022).

5 We acknowledge that the distinction between individual level and aggregate level can be viewed from different perspectives for calibration approaches where often also overall effects of user-individual calibration effects are reported, e.g., in the average reduction of the gap between the user profile and recommendation characteristics (Jugovac et al., 2017).

job but which also takes side constraints into account, for instance, the presumed willingness of a candidate to change positions. The authors leveraged a constraint-based optimization technique to solve that problem.

Technically, a variety of approaches to balance competing goals can be found in the literature. Reranking accuracy-optimized lists is probably the most common technique and was also used in early approaches for diversification in recommender systems (Bradley and Smyth, 2001). In this work, the particular goal was to diversify the recommendations returned by a *content-based* (case-based) system, which by design are similar to mostly non-diverse results. Notably, to quantify the diversity of a given list, the authors relied on a metric which was later on called *intra-list diversity* in the literature. Technically, three different diversification strategies (randomized, optimizing, greedy) were proposed and evaluated in their work. Generally, reranking techniques were applied in earlier information retrieval settings, in particular in the form of Maximal Marginal Relevance re-ranking (Carbonell and Goldstein, 1998). Since optimal re-ranking strategies are often computationally complex, heuristic or greedy approaches are more common in the literature, e.g., Adomavicius and Kwon (2012), Jugovac et al. (2017), and Abdollahpouri et al. (2021).

An alternative technical approach is taken in Jambor and Wang (2010), where the authors propose a framework based on *constrained linear optimization* to balance potentially competing optimization goals. Their framework primarily considers the assumed *utility* of an item for a given user (e.g., based on a predicted rating), but can also take additional constraints into account in the optimization process. Two example use cases are discussed, (a) promoting long-tail items and (b) the consideration of resources constraints, e.g., stock availability. Experimental evaluations indicate that balancing the trade-offs can be achieved with limited loss in accuracy. As in many other works, the main question however remains how to determine the right trade-off threshold in practice.

An optimization-based method was also proposed in Zhang and Hurley (2008), here with the objective of diversifying the recommendations through a side constraint while maintaining accuracy. The authors propose three ways of formulating the problem. One first possible objective of the optimization task was formulated as to maximize the diversity of the recommendation set while ensuring that the "matching value" (i.e., the preference match or utility for the user) does not fall beyond some tolerance value. An alternative formulation could be to maximize utility while reaching a certain level of diversity. Finally, a problem formulation with a combined optimization goal with a weighting parameter is possible as well. This last suggested problem formulation can be modeled as a binary quadratic programming problem with linear constraints, and the authors present a corresponding solution in their paper.

The *Auralist* framework proposed in Zhang et al. (2012) is designed to deliver not only relevant but also diversified and serendipitous music recommendations. Differently from optimization-based approaches, it works by combining the output of different ranking strategies: an accuracy-based one, one which promotes artists with diverse leadership, and one designed to help users break out of their personal music bubbles. A related approach of combining algorithms with different characteristics is proposed also in Ribeiro et al. (2015). In this latter work, an evolutionary algorithm is used to find a Pareto-efficient hybrid of the different algorithms. While the work in Ribeiro et al. (2015) is only assessed through offline experiments, the authors of *Auralist* evaluated their framework both offline and with the help of a user study. One key insight of the experiments is that serendipitous recommendations indeed lead to higher user satisfaction, despite a certain trade-off in accuracy that was observed in the offline experiments.

A comparison of offline results and a user-centric evaluation is also reported in Said et al. (2013). Here, the authors modified the traditional user-based nearest-neighbor method to consider the ratings of the most distant ("furthest") neighbors for the predictions. Offline experiments showed that this modification may lead to a notable performance drop in offline experiments. The user study however revealed that the modification did not negatively impact the perceived usefulness of the recommendations, even though they were very different in various dimensions (e.g., novelty, obviousness) than those provided by the traditional algorithm.

More sophisticated, graph-based algorithms for balancing accuracy and other factors, including diversity, were proposed in Zhou et al. (2010) and Isufi et al. (2021). In Zhou et al. (2010), a "heat-spreading" algorithm is applied to the graph formed based on the user-item interaction data. Like in several other works, the authors examine the trade-off between accuracy and other aspects through offline experiments[6]. Isufi et al. (2021) propose a graph convolution approach, building on ideas from Said et al. (2013) discussed earlier, and which only relies on rating information in the recommendation process. Again, offline experiments are conducted to study the accuracy-diversity (and coverage) trade-off.

An alternative technical approach to balance accuracy and novelty is put forward in de Souza Pereira Moreira et al. (2019). In this work, the authors present a generic meta-architecture for news recommendation problems, an application setting where the novelty of the items is often highly related to their relevance. Technically, the use of a parameterizable two-element loss function is proposed, where one part of the loss function targets accuracy and the other novelty. A streaming-based offline evaluation protocol is used to simulate real-world scenarios, and the effects of different hyperparameter settings for the loss function on the accuracy-novelty trade-off are studied.

Finally, McInerney et al. (2018) study the well-known explore-exploit dilemma in recommendation, where the system has the option to either recommend items of which it is relatively sure the user will like, or to take a more risky action and recommend items that should help the system to learn more about the user's preferences. In the latter case, exploring can be seen as taking a chance on an item with the hope that the user will actually like it. One possible problem when only exploiting is that the recommendations can be of limited novelty, and ultimately lead to limited user satisfaction in the long run. In their work, the authors study a contextual bandit approach in the music domain, which also involved the presentation of explanations to the users. Offline experiments on real-world logged interaction data and

6 While the authors claim to "solve" the accuracy-diversity dilemma, they technically propose specific measures to gauge the level of personalization and novelty of the recommendations. Their definition of diversity is not depending on item features and is thus rather uncommon in the literature.

a partially restricted A/B test provide solid indications for the practical usefulness of the approach.

2.2.3. Discussion

In many cases, optimizations performed at one level, such as the individual level, may affect the other level and vice versa. For example, when calibrating the recommendations for a user to match their *individual* diversity preferences, this will also be reflected to a certain extent on common diversity measures like intra-list diversity, when measured at the population-wide (*aggregate*) level. However, the relationship between individual-level and aggregate-level optimizations and the resulting effects may however not always be trivial in nature. Klimashevskaia et al. (2022), for example, found that calibrating recommendations with respect to popularity had a clear impact on the recommendations lists for *some users*, but it was found that the desired aggregate effect of reducing the popularity bias of the recommendations across users was not as substantial as expected. Similar considerations can be made for other quality objectives. For example, when optimizing recommendations for the individual user's *value-for-money* objective, this may have an impact on the overall revenue at the aggregate level. In sum, it therefore often seems advisable to observe multiple metrics in parallel to be able to understand the potentially subtle relationships between individual optimization goals.

2.3. Multistakeholder objectives

The beyond-accuracy quality metrics discussed in the previous section were historically mostly introduced to improve recommendations for end users. Higher diversity, for example, should avoid monotonicity, and novelty should support discovery. The underlying assumption—also of pure accuracy-oriented works—is that improving different quality aspects for users would be the sole factor for a successful recommender. Only in recent years, more attention has been paid in the literature to the fact that many recommendation scenarios in the real world are situated in environments, where the objectives of multiple stakeholders have to be considered. The common players in such *multistakeholder* recommendation problems include *end consumers*, the recommendation *platform*[7], item *providers* (suppliers), and sometimes even parts of a broader *society* (Abdollahpouri et al., 2020; Jannach and Bauer, 2020). In such settings, a recommender system may serve different purposes for different stakeholders (Jannach and Adomavicius, 2016), and the related objectives may stand in conflict.

In some cases there may even be subgroups within the consumer stakeholder group that have to be considered. These subgroups may have different expectations when using the service, and a recommender system should take these into account. In the music domain, for example, there can be different types of consumers, where one group's goal might lie in the exploration of the catalog and another group might be more interested in mood enhancement, see Bogt et al. (2011). The corresponding algorithms should then try to take the users' goals appropriately into account, see also Kapoor et al. (2015). Subgroups in a consumer stakeholder group can however also be identified by the providers, e.g., free vs. premium or new vs. existing customers, for which different objectives may exist. A number of recent research works in particular in the area of *fair recommender systems* address this latter problem. In Li et al. (2021) and Wu et al. (2022), for example, the authors investigate if highly active and less active users (including cold-start users) receive recommendations of largely different quality.

A typical problem setting in practice that involves *two stakeholders* is that of balancing consumer and platform objectives. In many cases, there may be a potential trade-off between (a) recommending the *most relevant* items for consumers and (b) recommending items that are also somewhat relevant but assumed to be favorable in terms of the platform's business objectives[8]. Some of the discussed beyond-accuracy metrics can actually be seen as serving both stakeholders. Making more novel recommendations not only potentially leads to a better user experience, but also to more engagement with the service and longer-term customer retention, which is an important platform goal in many application contexts (Anderson et al., 2020).

A number of research works however also consider monetary more directly, in particular in the form of recommender systems that are "price and profit aware." For example, Jannach and Adomavicius (2017) proposes a simple profit-aware recommendation approach *via* a simulation on a movie dataset by incorporating purchase-oriented information such as the price of the movie, sales probabilities, and the resulting profit, and shows that the approach can generate recommendations with yield higher profit with minimum loss in the relevance of the recommended movies. In Chen et al. (2008), as another work, two heuristic profit-aware strategies are proposed and the authors found that such methods can increase the profit from cross-selling without losing much recommendation accuracy.

Following a quite different technical approach, Wang and Wu (2009) develop an analytical model and optimization-based framework, which allows to *numerically study* the (short-term) effects of different marketing strategies. Possible strategies for example include a profit maximization approach or a "win-win" strategy for the platform and for consumers. The underlying model not only considers the relevance of the items that can be recommended to users, but also the items' selling price and profit. Moreover, budget constraints on the consumers' side are modeled as well. To address the challenges of fast online recommendation, an efficient solving strategy is proposed.

Differently from the works discussed so far, Azaria et al. (2013) investigate the effects of profit-aware and "value-aware" recommendation strategies through a *user study*. Two strategies are proposed which can be applied on top of any black-box recommendation model. In one strategy ("Hidden Agenda"), no prices for the items are present, whereas in the other ("Revenue

7 This is sometimes called the *service providers* in the literature.

8 See Shih and Kaufmann (2011) for a discussion of Netflix DVD recommendation strategy in 2011, which aimed to promote items that are less costly than blockbusters.

Maximizing") sales prices are considered. In the user study, participants received personalized recommendations and were then informed, among other aspects, about their satisfaction with the recommendation and their willingness to pay (WTP) for individual movies. The results show that the developed strategies can markedly increase the profit of the platform without a measurable drop in user satisfaction.

The results from a *field study* in the form of a *randomized controlled trial* are reported in Panniello et al. (2016). The specific goal of the study was to investigate the consumers' reactions in terms of purchasing behavior and (long-term) trust when confronted with recommendations that aim to balance accuracy and profitability. The experimental design included a profit-aware algorithm and a profit-agnostic one, and the recommendations were delivered to customers through personalized newsletters. The analyses after a 9-week period showed that higher profit can be achieved without a loss in consumer trust. Moreover, it turned out that the profit gains could be attributed to a combination of factors, consumer trust, diversity, and the relevance of the recommendations.

Besides situations with potential trade-offs at the recommendation platform side, there is the specific setting of *group recommendation*, a problem that has been studied for several years, even though not under the name multistakeholder recommendation (Masthoff, 2015). In such settings, the system's goal is to determine a set of recommendations that suit the preferences of a group of users, e.g., friends who want to watch a movie together. A unique aspect of such settings is that all involved (consumer) stakeholders in some ways receive or have to accept the same recommendation, which may or may not fit their preferences very well. A variety of strategies to aggregate individual user preferences were proposed over the years. Early works on the topic can be found in O'Connor et al. (2001) and Masthoff (2004). In Masthoff (2004), for instance, Masthoff reports the outcomes of different user studies aimed to understand how humans make choices for a group and find that humans indeed sometimes follow strategies inspired by Social Choice Theory (Sen, 1986). We iterate here that the group recommendation setting differs from other multistakeholder scenarios in that all stakeholders receive the same set of recommendations.

Reciprocal recommendation is another specific set of problem settings involving multiple stakeholders. Here, instead of recommending items to users, the problem is to recommend users to users, also known as people-to-people recommendation. Typical application scenarios are recommendations on dating (Pizzato et al., 2010) and recruiting platforms (Siting et al., 2012). A particularity of such settings is that the success of a recommendation is not determined solely by the recipient of the recommendation, but there must be a mutual preference match or compatibility between the two people involved, see Palomares et al. (2021) for an in-depth discussion on the topic. The recommendation platform (service provider), therefore, faces additional complexities in the matching process and in parallel has to observe its own business objectives and constraints. On a job recommendation platform, for example, the platform may have to additionally ensure that each paid job advertisement receives a minimum number of relevant impressions, i.e., exposure (Abel et al., 2017).

Similar considerations may generally apply when the recommendation platform serves as a marketplace with multiple suppliers of identical or comparable items. Let us consider again the example of a typical hotel booking platform, which serves personalized recommendations to its users (Jannach and Bauer, 2020). Besides the consumer, who already might have competing objectives, there are the property owners, who have their offerings listed on the booking platform and pay a commission for each booking. The goal of the property owners is that their offerings are exposed to as many matching customers as possible in order to increase the chances of being booked. The booking platform, finally, may not only be interested in recommending matching hotels to consumers but might also seek to maximize their commission, e.g., by recommending slightly more expensive hotels. In addition, to balance these objectives, the platform may furthermore have to ensure that *all* listed properties reach a sufficient level of exposure, i.e., chance of being booked. This may be required to ensure a long-term relationship with property owners, who might otherwise discontinue listing their offerings on the platform at some stage (Krasnodebski and Dines, 2016).

2.4. Time horizon objectives

In some application domains, it might be quite simple to increase short-term Key Performance Indicators. In the hotel booking scenario which we have just discussed, boosting short-term revenue might be achieved by recommending hotels with currently discounted rates, which maximizes the probability of a transaction (Jannach et al., 2017). In the news domain, recommending articles on trending topics, articles with click-bait headlines, or generally popular content such as celebrity gossip may lead to high click-through rates (CTR). In the music domain, recommending tracks of trending or popular artists, which the user already knows, might be a safe strategy when the target metric is to avoid "skip" events.

Such strategies that are successful in the short term may however be non-optimal or even detrimental in the long run. The recommendation of discounted hotel rooms may be bad for profit, and recommending hotels that lead to the highest commission may hurt consumer trust. News readers may be disappointed when actually reading articles with a click-bait headline and may not trust these recommendations in the future. Music listeners finally may have difficulties discovering new artists over time and may quit using the service after some time.

Most academic research is based on one-shot evaluations, typically focusing on prediction accuracy given a static dataset and a certain point in time. The longitudinal effects of different recommendation strategies are much less explored and there is also limited literature on the long-term effects of recommender systems in the industry. A/B tests in the industry may last from a few weeks to several months. In Gomez-Uribe and Hunt (2015), the case of Netflix is discussed, where one main KPI is customer retention, which is oriented toward the long-term perspective. In their case, attributing changes in the recommender system to such long-term effects is reported to be challenging, e.g., because of already high retention rates and the need for large user samples. Other reports from real-world deployments of recommender systems can be found in Panniello et al. (2016) or Lee and Hosanagar (2019). In Lee and Hosanagar (2019), the authors for example found that using a recommender system led to decreased sales diversity compared to a situation without a recommender.[9] A similar effect was reported in Anderson et al. (2020), where the recommender system on a music streaming site led to a reduced aggregate consumption diversity. A survey of other reports on real-world applications of recommender systems can be found in Jannach and Jugovac (2019).

Given the limitations of one-shot evaluations, we have observed an increased interest in longitudinal studies in recent years. One prominent line of research lies in the area of reinforcement learning (RL) approaches in particular in the form of contextual bandits, see e.g., Li et al. (2010) for earlier work in the news domain. In such approaches, the system sequentially selects items to recommend to users and then incorporates the users' feedback for subsequent recommendations. Different recommendation algorithms can be evaluated offline with the help of simulators, e.g., Rohde et al. (2018) and McInerney et al. (2021). A common challenge in this context is to ensure that such evaluations are unbiased (Li et al., 2011; Huang et al., 2020).[10] We note that the consideration of temporal aspects such as different time horizons or delayed feedback has been explored in the RL literature for the related problem of computational advertising for several years (Chapelle, 2014; Theocharous et al., 2015).

Reinforcement learning approaches typically aim at finding a strategy to maximize the expected *reward*. During the last few years, a number of studies that use *other forms* of simulations were published that focus on other important long-term phenomena of recommender systems. These studies for example focus on longitudinal effects of recommender systems on sales diversity (Fleder and Hosanagar, 2009), potential reinforcement effects in terms of popularity bias, and other aspects for traditional and session-based recommendations (Jannach et al., 2015b; Ferraro et al., 2020), longitudinal performance effects of recommender systems and the "performance paradox" (Zhang et al., 2019), differences in terms of long-term effects of consumer-oriented and profit-oriented recommendation strategies (Ghanem et al., 2022).

Directly optimizing for long-term rewards is typically hard due to the sparsity in observing these events and the low signal-to-noise ratio (weak connection) between these long-term outcomes and a single recommendation. Therefore, researchers often leverage surrogates or mid-level outcomes that are easier to observe as a proxy for potential long-term outcomes. For example, Wang et al. (2022) investigates several surrogates such as diversity of consumption, frequency of returning to the platform, repeated consumption, etc., as a proxy to estimate long-term user engagement. The authors then use such surrogates in the objective function for the RL algorithm to optimize for those metrics. With their work, they aim at providing guidance for researchers and practitioners when selecting surrogate measures to address the difficult problem of optimizing for long-term objectives.

9 It is worth noting that the authors studied one particular class of non-personalized recommendation algorithms here based on co-purchasing statistics ("Customers who bought this item also bought ...").

10 A critical discussion of current evaluation practices when applying RL for sequential problems can be found in Deffayet et al. (2022).

2.5. User experience objectives

Going beyond the specifics of individual algorithms, there can be also various objectives to be pursued at the user interaction level of a recommender system. The design space for the user interface of recommender systems is actually large, see Jugovac and Jannach (2017), and there thus may be a number of competing objectives at the user interface (UI) level.

Here, we only list a few examples of potential trade-offs that may be common for many recommender system applications.

- *Information completeness vs. information overload*: This, for instance, refers to the question of how many items should be shown to users and if we should completely filter out certain items from the result list. Showing too few options may give users the feeling that the system holds back some information. If there is too much information users will find themselves again in a situation of information overload (Bollen et al., 2010; Aljukhadar et al., 2012). Besides the question of how many options to show, a related question is how much detail and additional information to show for each recommendation.
- *Transparency and user control vs. cognitive effort*: Transparency and explanations are commonly considered to be trust-establishing factors in recommender systems (Pu et al., 2011). A variety of different ways of explaining recommendations were proposed in the literature (Tintarev and Masthoff, 2011; Nunes and Jannach, 2017). Many of these academic proposals are quite complex and may easily cognitively overload average end users. Similar considerations apply for approaches that implement mechanisms for *user control* in recommender systems (Ekstrand et al., 2015; Jannach et al., 2016).
- *Flexibility vs. efficiency*: This question arises in the context of modern conversational recommender systems that are implemented in the form of chatbots. Chatbots typically support two forms of interactions: a) natural language input and b) form-based input (i.e., using buttons). While natural language inputs may allow for more flexible interactions, the study in Iovine et al. (2020), for instance, indicated that a combination of interaction modalities was most effective.

Several other more general design trade-offs may exist depending on the specific application, e.g., regarding acceptable levels of automating adaptivity of the user interface, which may hamper usability (Paymans et al., 2004).

2.6. Engineering objectives

In this final category, we discuss technical aspects and their potential trade-offs. We call them "engineering objectives", as they refer to more general system properties.

One such trade-off in practice may lie in the complexity of the underlying algorithms and the gains that one may obtain in terms of business-related KPIs. Already in the context of the Netflix Prize (Bennett and Lanning, 2007) we could observe that the winning solutions were finally not put into production, partly due to their complexity. Similar considerations can be made for today's sometimes computationally demanding methods based on deep learning. In some cases, there might be a diminishing return on deploying the most sophisticated models in production, only because they lead to slightly better accuracy values in offline testing. In some research works, it even turns out that "embarrassingly shallow" models can be highly competitive in offline evaluations (Steck, 2019).

With highly complex models, not only scalability issues may arise and monetary costs for computing resources may increase, but the complexity of the architectures might also make such systems more difficult to maintain, debug, and explain. On the other hand, solutions built upon modern deep learning frameworks are sometimes reported to be advantageous over conceptually simpler, but specialized solutions, because these frameworks and deep learning architectures make it very easy to integrate various types of information into the models (Steck et al., 2021).

However, integrating different types of information can also come at a price. In many organizations, the different pieces of information that should be integrated into a recommender system—e.g., user behavior logs, purchase records, item meta-data, stock availability, and business rules—may be stored in various systems and databases. This can make data integration and data quality assurance a highly challenging task, in cases where increasingly more data sources must be combined.

3. Summary and challenges

Our review outlines that providing automated recommendations is a problem that may require the consideration of more than one objective in many real-world use cases. Such multi-objective settings may include competing objectives of consumers, possible tensions between the goals of different stakeholders, conflicts when optimizing for different time horizons, competing design choices at the UI level, as well as system-level and engineering-related considerations. In this work, we reviewed the literature in this area and provided a taxonomy to organize the various dimensions of multi-objective recommendation. We note here that the categories of the taxonomy are not mutually exclusive. For instance, a multi-objective recommendation approach may address both aspects regarding different time horizons as well as the possibly competing goals of the involved stakeholders.

In practice, one main challenge may usually lie in deciding on the right balance between the competing goals from an organizational perspective. Various stakeholders from different organizational units may have to agree on such decisions, and corresponding KPIs need to be defined and monitored. Given these KPIs, suitable optimization goals and possibly proxy measures have to be implemented and validated at the technical level.

In academic settings, researchers typically abstract from the specifics of a given application context, aimed at developing generalizable algorithmic solutions to deal with multi-objective problem settings. This abstraction process commonly involves the use of *offline evaluation approaches*, the establishment of certain assumptions, and the introduction of computational metrics which should be optimized. After such an abstraction, one main challenge, however, lies in the evaluation process and, in particular, in making sure that improvements that are observed in terms of abstract evaluation measures would translate to better systems in practice (Cremonesi and Jannach, 2021).

Unfortunately, in many of today's research works, we observe phenomena similar to the "abstraction traps" described by Selbst et al. (2019) in the context of research on algorithmic works in *Fair Machine Learning*. In the case of competing individual-level quality goals, for example, how can we be sure that a particular diversity metric, which we optimize such as an intra-list similarity, matches human perceptions and what would be the right balance for a given application setting or an individual user? How do we know if calibrated recommendations are liked more by users, and what would be the effects of calibration on organizational goals? Answering such questions requires corresponding user studies to, e.g., validate that the computational metrics are good proxies for human perceptions. An attempt to investigate the relationship between *perceived diversity* and the widely used intra-list similarity measure can be found in Jesse et al. (2022).

The problem however becomes even more challenging when not even the target concepts are entirely clear. In recent years, a widely investigated multi-objective problem setting is the provision of *fair* recommendations (Ekstrand et al., 2022). Unfortunately, optimizing for fairness turns out to be challenging, as fairness is a societal construct, and a number of definitions exist, see Narayanan (2018). Researchers in computer science, therefore, came up with various types of ways of operationalizing fairness constraints. However, in many of such works, little or no evidence or argumentation is provided why the chosen fairness metrics are meaningful in practice in general or in a particular application setting, see Deldjoo et al. (2022) for a survey on the recent literature.

In some cases, including our own previous work, e.g., Abdollahpouri et al. (2019a), making fair recommendations is only loosely connected or even simply equated with reducing the popularity bias of recommendations. Technically, this is often done by matching it with a target distribution or metric threshold, which is assumed to be given. In reality, however, it is not clear what would be the underlying *normative claim* that mandates that less popular items should be recommended. In fact, many of these unpopular items might simply be of poor quality. Moreover, users might not even *perceive* such recommendations of unpopular items to be fair. However, there are also studies that indicate that recommending mostly popular items may negatively impact accuracy, and, importantly, that these effects may differ across user groups. Our previous study in the movie domain (Abdollahpouri et al., 2019b), for example, indicated that users of the group with the least mainstream taste received the worst recommendations. A similar observation was later made in the music domain by Kowald et al. (2020). We note that here, item popularity is often assessed by counting the number of past interactions in the database. The assumed fairness problem is thus related, but different from the *item cold-start* problem (Panda and Ray, 2022). Recommending such items is of course important in practice, to ensure a certain level of initial exposure to new items.

Overall, these observations call for more studies involving humans in the evaluation loop and industry partners in the research process. However, only a few works exist in that direction so far. An example of a user study can be found in Azaria et al. (2013), and outcomes of field studies are described in Panniello et al. (2016). An offline evaluation with real-world data from the industry is done in Mehrotra et al. (2018), but even in this case, it is not clear if the computational metrics truly correspond to the real-world goals, e.g., if more listening events on the music platform lead to higher user satisfaction as claimed.

Ultimately, despite such recent progress, multi-objective recommender systems remains a highly important research area with a number of challenging research questions. Addressing such questions will however help to pave the way toward more impactful recommender systems research in the future.

Author contributions

DJ and HA: conceptualization, research, and writing. All authors contributed to the article and approved the submitted version.

Acknowledgments

The content of this manuscript has been presented in part at the 2nd Workshop on Multi-Objective Recommender Systems co-located with the ACM Conference on Recommender Systems 2022, Seattle, USA (Jannach, 2022).

Conflict of interest

HA is employed by Spotify Inc.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdollahpouri, H., Adomavicius, G., Burke, R., Guy, I., Jannach, D., Kamishima, T., et al. (2020). Multistakeholder recommendation: survey and research directions. *User Model Useradapt Interact.* 30, 127–158. doi: 10.1007/s11257-019-09256-1

Abdollahpouri, H., and Burke, R. (2022). "Multistakeholder recommender systems," in *Recommender Systems Handbook*, eds F. Ricci, L. Rokach, and B. Shapira (New York, NY: Springer U.S.), 647–677.

Abdollahpouri, H., Burke, R., and Mobasher, B. (2017). "Controlling popularity bias in learning-to-rank recommendation," in *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys '17* (Como), 42–46.

Abdollahpouri, H., Burke, R., and Mobasher, B. (2019a). "Managing popularity bias in recommender systems with personalized re-ranking," in *Proceedings of the Thirty-Second International Florida Artificial Intelligence Research Society Conference (FLAIRS '19)* (Sarasota, FL), 413–418.

Abdollahpouri, H., Mansoury, M., Burke, R., and Mobasher, B. (2019b). The unfairness of popularity bias in recommendation. *arXiv preprint* arXiv:1907.13286. doi: 10.48550/arXiv.1907.13286

Abdollahpouri, H., Mansoury, M., Burke, R., Mobasher, B., and Malthouse, E. C. (2021). "User-centered evaluation of popularity bias in recommender systems," in *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2021* (Barcelona), 119–129.

Abel, F., Deldjoo, Y., Elahi, M., and Kohlsdorf, D. (2017). "Recsys challenge 2017: offline and online evaluation," in *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys 2017* (Como), 372–373.

Adomavicius, G., and Kwon, Y. (2007). New recommendation techniques for multicriteria rating systems. *IEEE Intell. Syst.* 22, 48–55. doi: 10.1109/MIS.2007.58

Adomavicius, G., and Kwon, Y. (2012). Improving aggregate recommendation diversity using ranking-based techniques. *IEEE TKDE* 24, 896–911. doi: 10.1109/TKDE.2011.15

Adomavicius, G., and Kwon, Y. (2015). "Multi-criteria recommender systems," in *Recommender Systems Handbook*, eds F. Ricci, L. Rokach, and B. Shapira (New York, NY: Springer U.S.), 847–880.

Aljukhadar, M., Senecal, S., and Daoust, C.-E. (2012). Using recommendation agents to cope with information overload. *Int. J. Electr. Commerce* 17, 41–70. doi: 10.2753/JEC1086-4415170202

Anderson, A., Maystre, L., Anderson, I., Mehrotra, R., and Lalmas, M. (2020). "Algorithmic effects on the diversity of consumption on spotify," in *Proceedings of The Web Conference 2020, WWW '20* (Taipei), 2155–2165.

Azaria, A., Hassidim, A., Kraus, S., Eshkol, A., Weintraub, O., and Netanely, I. (2013). "Movie recommender system for profit maximization," in *Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13* (Hong Kong), 121–128.

Bennett, J., and Lanning, S. (2007). "The Netflix prize," in *Proceedings of KDD Cup and Workshop, Vol. 2007* (San Jose, CA), 35.

Bogt, T. F. T., Mulder, J., Raaijmakers, Q. A., and Gabhainn, S. N. (2011). Moved by music: a typology of music listeners. *Psychol. Music* 39, 147–163. doi: 10.1177/0305735610370223

Bollen, D. G. F. M., Knijnenburg, B. P., Willemsen, M. C., and Graus, M. P. (2010). "Understanding choice overload in recommender systems," in *Proceedings of the 4th Conference on Recommender Systems (RecSys '10)* (Barcelona), 63–70.

Bradley, K., and Smyth, B. (2001). "Improving recommendation diversity," in *Proceedings of the 12th National Conference in Artificial Intelligence and Cognitive Science* (Maynooth), 75–84.

Burke, R. D., Hammond, K. J., and Yound, B. (1997). The FindMe approach to assisted browsing. *IEEE Expert.* 12, 32–40. doi: 10.1109/64.608186

Cacheda, F., Carneiro, V., Fernández, D., and Formoso, V. (2011). Comparison of collaborative filtering algorithms: limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Trans. Web* 5, 1–33. doi: 10.1145/1921591.1921593

Carbonell, J., and Goldstein, J. (1998). "The use of mmr, diversity-based reranking for reordering documents and producing summaries," in *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '98* (Melbourne), 335–336.

Chapelle, O. (2014). "Modeling delayed feedback in display advertising," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14* (New York, NY), 1097–1105.

Chen, L., and Pu, P. (2012). Critiquing-based recommenders: survey and emerging trends. *User Model Useradapt Interact.* 22, 125–150. doi: 10.1007/s11257-011-9108-6

Chen, L.-S., Hsu, F.-H., Chen, M.-C., and Hsu, Y.-C. (2008). Developing recommender systems with the consideration of product profitability for sellers. *Inf. Sci.* 178, 1032–1048. doi: 10.1016/j.ins.2007.09.027

Cremonesi, P., and Jannach, D. (2021). Progress in recommender systems research: Crisis? *what crisis? AI Mag.* 42, 43–54. doi: 10.1609/aimag.v42i3.18145

de Souza Pereira Moreira, G., Jannach, D., and da Cunha, A. M. (2019). Contextual hybrid session-based news recommendation with recurrent neural networks. *IEEE Access* 7, 169185–169203. doi: 10.1109/ACCESS.2019.2954957

Deffayet, R., Thonet, T., Renders, J.-M., and de Rijke, M. (2022). Offline evaluation for reinforcement learning-based recommendation: a critical issue and some alternatives. *SIGIR Forum* 56, 1–14. doi: 10.1145/3582900.3582905

Deldjoo, Y., Jannach, D., Bellogin, A., Difonzo, A., and Zanzonelli, D. (2022). A survey of research on fair recommender systems. *arXiv:2205.11127 [cs.IR]*. doi: 10.48550/arXiv.2205.11127

Ekstrand, M. D., Das, A., Burke, R., and Diaz, F. (2022). Fairness in information access systems. *Foundat. Trends® Inf. Retrieval* 16, 1–177. doi: 10.1561/9781638280415

Ekstrand, M. D., Kluver, D., Harper, F. M., and Konstan, J. A. (2015). "Letting users choose recommender algorithms: |an experimental study," in *Proceedings of the 9th ACM Conference on Recommender Systems, RecSys '15* (Vienna), 11–18.

Felfernig, A., Friedrich, G., Jannach, D., and Zanker, M. (2015). "Constraint-based recommender systems," in *Recommender Systems Handbook* (New York, NY: Springer), 161–190.

Ferraro, A., Jannach, D., and Serra, X. (2020). "Exploring longitudinal effects of session-based recommendations," in *Proceedings of the 2020 ACM Conference on Recommender Systems*, 474–479.

Fleder, D., and Hosanagar, K. (2009). Blockbuster culture's next rise or fall: the impact of recommender systems on sales diversity. *Manag. Sci.* 55, 697–712. doi: 10.1287/mnsc.1080.0974

Gao, C., Lei, W., He, X., de Rijke, M., and Chua, T.-S. (2021). Advances and challenges in conversational recommender systems: a survey. *AI Open* 2, 100–126. doi: 10.1016/j.aiopen.2021.06.002

Ghanem, N., Leitner, S., and Jannach, D. (2022). Balancing consumer and business value of recommender systems: a simulation-based analysis. *Electron. Commerce Res. Appl.* 55, 101195. doi: 10.1016/j.elerap.2022.101195

Gomez-Uribe, C. A., and Hunt, N. (2015). The Netflix recommender system: algorithms, business value, and innovation. *Trans. Manag. Inf. Syst.* 6, 1–19. doi: 10.1145/2843948

Gunawardana, A., and Shani, G. (2015). "Evaluating recommender systems," in *Recommender Systems Handbook*, eds F. Ricci, L. Rokach, and B. Shapira (New York, NY: Springer), 265–308.

He, C., Parra, D., and Verbert, K. (2016). Interactive recommender systems: a survey of the state of the art and future research challenges and opportunities. *Expert. Syst. Appl.* 56, 9–27. doi: 10.1016/j.eswa.2016.02.013

Herlocker, J. L., Konstan, J. A., and Riedl, J. (2000). "Explaining collaborative filtering recommendations," in *CSCW '00* (Minneapolis, MN), 241–250.

Huang, J., Oosterhuis, H., de Rijke, M., and van Hoof, H. (2020). "Keeping dataset biases out of the simulation: a debiased simulator for reinforcement learning based recommender systems," in *Fourteenth ACM Conference on Recommender Systems, RecSys* '20, 190–199.

Huang, S. (2011). Designing utility-based recommender systems for e-commerce: evaluation of preference-elicitation methods. *Electron Commer. Res. Appl.* 10, 398–407. doi: 10.1016/j.elerap.2010.11.003

Iovine, A., Narducci, F., and Semeraro, G. (2020). Conversational recommender systems and natural language: a study through the ConveRSE framework. *Decis. Support Syst.* 131, 113250–113260. doi: 10.1016/j.dss.2020.113250

Isufi, E., Pocchiari, M., and Hanjalic, A. (2021). Accuracy-diversity trade-off in recommender systems via graph convolutions. *Inf. Process. Manag.* 58, 102459. doi: 10.1016/j.ipm.2020.102459

Jambor, T., and Wang, J. (2010). "Optimizing multiple objectives in collaborative filtering," in *Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys* '10 (Barcelona), 55–62.

Jannach, D. (2004). "ADVISOR SUITE-A knowledge-based sales advisory system," in *ECAI '04* (Valencia), 720–724.

Jannach, D. (2022). "Multi-objective recommendation: overview and challenges," in *Proceedings of the 2nd Workshop on Multi-Objective Recommender Systems co-located with 16th ACM Conference on Recommender Systems (RecSys 2022), volume 3268 of CEUR Workshop Proceedings* (Seattle). Available online at: https://ceur-ws.org/Vol-3268/

Jannach, D., and Adomavicius, G. (2016). "Recommendations with a purpose," in *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16* (Boston, MA), 7–10.

Jannach, D., and Adomavicius, G. (2017). Price and profit awareness in recommender systems. *CoRR abs/1707.08029*. doi: 10.48550/arXiv.1707.08029

Jannach, D., and Bauer, C. (2020). Escaping the McNamara fallacy: towards more impactful recommender systems research. *AI Mag.* 41, 79–95. doi: 10.1609/aimag.v41i4.5312

Jannach, D., and Jugovac, M. (2019). Measuring the business value of recommender systems. *ACM TMIS* 10, 1–23. doi: 10.1145/3370082

Jannach, D., Karakaya, Z., and Gedikli, F. (2012). "Accuracy improvements for multi-criteria recommender systems," in *Proceedings of the 13th ACM Conference on Electronic Commerce (ACM EC 2012)* (Valencia), 674–689.

Jannach, D., Lerche, L., and Kamehkhosh, I. (2015a). "Beyond "hitting the hits"-generating coherent music playlist continuations with the right tracks," in *Proceedings of the 9th ACM Conference on Recommender Systems (RecSys 2015)* (Vienna), 187–194.

Jannach, D., Lerche, L., Kamehkhosh, I., and Jugovac, M. (2015b). What recommenders recommend: an analysis of recommendation biases and possible countermeasures. *User Model Useradapt Interact.* 25, 427–491. doi: 10.1007/s11257-015-9165-3

Jannach, D., Ludewig, M., and Lerche, L. (2017). Session-based item recommendation in e-commerce: on short-term intents, reminders, trends and discounts. *User Model Useradapt Interact.* 27, 351–392. doi: 10.1007/s11257-017-9194-1

Jannach, D., Manzoor, A., Cai, W., and Chen, L. (2021). A survey on conversational recommender systems. *ACM Comput. Surveys* 54, 1–36. doi: 10.1145/3453154

Jannach, D., Naveed, S., and Jugovac, M. (2016). "User control in recommender systems: overview and interaction challenges," in *17th International Conference on Electronic Commerce and Web Technologies (EC-Web 2016)* (Porto).

Jannach, D., and Zanker, M. (2021). "Impact and value of recommender systems," in *Recommender Systems Handbook*, eds F. Ricci, B. Shapira, and L. Rokach (New York, NY: Springer U.S.).

Jesse, M., Bauer, C., and Jannach, D. (2022). Intra-list similarity and human diversity perceptions of recommendations: The details matter. *User Model. User Adapted Interact*. doi: 10.1007/s11257-022-09351-w

Jugovac, M., and Jannach, D. (2017). "Interacting with recommenders - overview and research directions," in *ACM Transactions on Intelligent Interactive Systems (ACM TiiS)*, 7, 1–46.

Jugovac, M., Jannach, D., and Lerche, L. (2017). Efficient optimization of multiple recommendation quality factors according to individual user tendencies. *Expert. Syst. Appl.* 81, 321–331. doi: 10.1016/j.eswa.2017.03.055

Kaminskas, M., and Bridge, D. (2016). Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Trans. Interact. Intell. Syst. 7* 2, 42. doi: 10.1145/2926720

Kamishima, T., Akaho, S., Asoh, H., and Sakuma, J. (2018). "Recommendation independence," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Vol. 81 of Proceedings of Machine Learning Research* (Nice), 187–201.

Kapoor, K., Kumar, V., Terveen, L., Konstan, J. A., and Schrater, P. (2015). "I like to explore sometimes": adapting to dynamic user novelty preferences," in *Proceedings of RecSys '15*, 19–26.

Klimashevskaia, A., Elahi, M., Jannach, D., Trattner, C., and Skjærven, L. (2022). "Mitigating popularity bias in recommendation: potential and limits of calibration approaches," in *Advances in Bias and Fairness in Information Retrieval*, eds L. Boratto, S. Faralli, M. Marras, and G. Stilo, 82–90.

Kowald, D., Schedl, M., and Lex, E. (2020). "The unfairness of popularity bias in music recommendation: a reproducibility study," in *Advances in Information Retrieval-42nd European Conference on IR Research, ECIR 2020*, 35–42.

Krasnodebski, J., and Dines, J. (2016). "Considering supplier relations and monetization in designing recommendation systems," in *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16* (Boston, MA), 381–382.

Lee, D., and Hosanagar, K. (2019). How do recommender systems affect sales diversity? A cross-category investigation via randomized field experiment. *Inf. Syst. Res.* 30, 239–259. doi: 10.1287/isre.2018.0800

Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). "A contextual-bandit approach to personalized news article recommendation," in *Proceedings of the 19th International Conference on World Wide Web, WWW '10* (Raleigh), 661–670.

Li, L., Chu, W., Langford, J., and Wang, X. (2011). "Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms," in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM '11* (Hong Kong), 297–306.

Li, R., Kahou, S. E., Schulz, H., Michalski, V., Charlin, L., and Pal, C. (2018). "Towards deep conversational recommendations," in *NIPS '18* (Montreal, QC), 9725–9735.

Li, Y., Chen, H., Fu, Z., Ge, Y., and Zhang, Y. (2021). User-oriented fairness in recommendation. *Proc. Web Conf.* 2021, 624–632. doi: 10.1145/3442381.3449866

Liu, D.-R., and Shih, Y.-Y. (2005). Integrating ahp and data mining for product recommendation based on customer lifetime value. *Inf. Manag.* 42, 387–400. doi: 10.1016/j.im.2004.01.008

Manouselis, N., and Costopoulou, C. I. (2007). Analysis and classification of multi-criteria recommender systems. *World Wide Web* 10, 415–441. doi: 10.1007/s11280-007-0019-8

Masthoff, J. (2004). "Group modeling: selecting a sequence of television items to suit a group of viewers," in *Personalized Digital Television: Targeting Programs to individual Viewers* (Netherlands: Springer), 93–141.

Masthoff, J. (2015). "Group recommender systems: Aggregation, satisfaction and group attributes," in *Recommender Systems Handbook*, eds F. Ricci, L. Rokach, and B. Shapira (New York, NY: Springer).

McInerney, J., Elahi, E., Basilico, J., Raimond, Y., and Jebara, T. (2021). "Accordion: a trainable simulator for long-term interactive systems," in *Fifteenth ACM Conference on Recommender Systems, RecSys* '21 (Amsterdam, NL), 102–113.

McInerney, J., Lacker, B., Hansen, S., Higley, K., Bouchard, H., Gruson, A., et al. (2018). "Explore, exploit, and explain: Personalizing explainable recommendations with bandits," in *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys '18* (Vancouver, CA), 31–39.

McNee, S. M., Riedl, J., and Konstan, J. A. (2006). "Being accurate is not enough: how accuracy metrics have hurt recommender systems," in *CHI '06 Extended Abstracts on Human Factors in Computing Systems, CHI EA '06* (Montreal, QC), 1097–1101.

Mehrotra, R., McInerney, J., Bouchard, H., Lalmas, M., and Diaz, F. (2018). "Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness satisfaction in recommendation systems," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM '18* (Turin), 2243–2251.

Narayanan, A. (2018). "Translation tutorial: 21 fairness definitions and their politics," in *Proceedings of Conferences Fairness Accountability Transparency,Vol. 1170* (New York, NY), 3.

Nunes, I., and Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Model. User-adapted Interact.* 27, 393–444. doi: 10.1007/s11257-017-9195-0

O'Connor, M., Cosley, D., Konstan, J. A., and Riedl, J. (2001). "Polylens: a recommender system for groups of users," in *ECSCW 2001: Proceedings of the Seventh European Conference on Computer Supported Cooperative Work* (Netherlands: Springer), 199–218.

Oh, J., Park, S., Yu, H., Song, M., and Park, S.-T. (2011). "Novel recommendation based on personal popularity tendency," in *ICDM* '11 (Vancouver, CA), 507–516.

Palomares, I., Porcel, C., Pizzato, L., Guy, I., and Herrera-Viedma, E. (2021). Reciprocal recommender systems: analysis of state-of-art literature, challenges and opportunities towards social recommendation. *Inf. Fusion* 69, 103–127. doi: 10.1016/j.inffus.2020.12.001

Panda, D. K., and Ray, S. (2022). Approaches and algorithms to mitigate cold start problems in recommender systems: a systematic literature review. *J. Intell. Inf. Syst.* 59, 341–366. doi: 10.1007/s10844-022-00698-5

Panniello, U., Hill, S., and Gorgoglione, M. (2016). The impact of profit incentives on the relevance of online recommendations. *Electron. Commer. Rec. Appl.* 20, 87–104. doi: 10.1016/j.elerap.2016.10.003

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding From You*. London: Penguin Group.

Paymans, T. F., Lindenberg, J., and Neerincx, M. (2004). "Usability trade-offs for adaptive user interfaces: ease of use and learnability," in *Proceedings of the 9th International Conference on Intelligent User Interfaces, IUI '04* (Madeira), 301–303.

Pizzato, L., Rej, T., Chung, T., Koprinska, I., and Kay, J. (2010). "Recon: a reciprocal recommender for online dating," in *Proceedings of the Fourth ACM Conference on Recommender Systems* (Barcelona), 207–214.

Pu, P., Chen, L., and Hu, R. (2011). "A user-centric evaluation framework for recommender systems," in *RecSys* '11, 157–164.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). "GroupLens: an open architecture for collaborative filtering of netnews," in *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, CSCW '94* (Chapel Hill), 175–186.

Ribeiro, M. T., Ziviani, N., Moura, E. S. D., Hata, I., Lacerda, A., and Veloso, A. (2015). Multiobjective pareto-efficient approaches for recommender systems. *ACM Trans. Intell. Syst. Technol.* 5, 350. doi: 10.1145/2629350

Rodriguez, M., Posse, C., and Zhang, E. (2012). "Multiple objective optimization in recommender systems," in *Proceedings of the Sixth ACM Conference on Recommender Systems, RecSys* '12 (Dublin), 11–18.

Rohde, D., Bonner, S., Dunlop, T., Vasile, F., and Karatzoglou, A. (2018). Recogym: a reinforcement learning environment for the problem of product recommendation in online advertising. *arXiv preprint* arXiv:1808.00720. doi: 10.48550/arXiv.1808.00720

Said, A., Fields, B., Jain, B. J., and Albayrak, S. (2013). "User-centric evaluation of a k-furthest neighbor collaborative filtering recommender algorithm," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW '13* (San Antonio), 1399–1408.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., and Lehman, D. R. (2002). Maximizing versus satisficing: happiness is a matter of choice. *J. Pers. Soc. Psychol.* 83, 1178–1197. doi: 10.1037/0022-3514.83.5.1178

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., and Vertesi, J. (2019). "Fairness and abstraction in sociotechnical systems," in *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT** '19 (Atlanta), 59–68.

Sen, A. (1986). Social choice theory. *Handbook Math. Econ.* 3, 1073–1181. doi: 10.1016/S1573-4382(86)03004-7

Shardanand, U., and Maes, P. (1995). "Social information filtering: algorithms for automating "word of mouth"," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI* '95 (Denver), 210–217.

Shi, L. (2013). "Trading-off among accuracy, similarity, diversity, and long-tail: a graph-based recommendation approach," in *Proceedings of the 7th ACM Conference on Recommender Systems* (Hong Kong), 57–64.

Shih, W., and Kaufmann, S. (2011). *Netflix in 2011*. Boston, MA: Harvard Business School.

Siting, Z., Wenxing, H., Ning, Z., and Fan, Y. (2012). "Job recommender systems: a survey," in *2012 7th International Conference on Computer Science Education (ICCSE)* (IEEE), 920–924.

Smyth, B. (2007). "Case-based recommendation," in *The Adaptive Web: Methods and Strategies of Web Personalization*, eds P. Brusilovsky, A. Kobsa, and W. Nejdl, 342–376.

Steck, H. (2018). "Calibrated recommendations," in *ACM RecSys* '18 (Vancouver, CA), 154–162.

Steck, H. (2019). "Embarrassingly shallow autoencoders for sparse data," in *The World Wide Web Conference, WWW* '19 (San Francisco, CA), 3251–3257.

Steck, H., Baltrunas, L., Elahi, E., Liang, D., Raimond, Y., and Basilico, J. (2021). Deep learning for recommender systems: a Netflix case study. *AI Mag.* 42, 7–18. doi: 10.1609/aimag.v42i3.18140

Theocharous, G., Thomas, P. S., and Ghavamzadeh, M. (2015). "Personalized ad recommendation systems for life-time value optimization with guarantees," in *Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15* (Buones Aires), 1806–1812.

Tintarev, N., and Masthoff, J. (2011). "Designing and evaluating explanations for recommender systems," in *Recommender Systems Handbook* (New York, NY: Springer), 479–510.

Triantaphyllou, E. (2000). "Multi-criteria decision making methods," in *Multi-criteria Decision Making Methods: A Comparative Study* (New York, NY: Springer), 5–21.

Vargas, S., and Castells, P. (2011). "Rank and relevance in novelty and diversity metrics for recommender systems," in *Proceedings of the Fifth ACM Conference on Recommender Systems* (Chicago, IL), 109–116.

Wang, H.-F., and Wu, C.-T. (2009). A mathematical model for product selection strategies in a recommender system. *Expert Syst. Appl.* 36(3, Part 2), 7299–7308. doi: 10.1016/j.eswa.2008.09.006

Wang, Y., Sharma, M., Xu, C., Badam, S., Sun, Q., Richardson, L., et al. (2022). "Surrogate for long-term user experience in recommender systems," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '22* (Washington, DC), 4100–4109.

Wu, C., Wu, F., Qi, T., and Huang, Y. (2022). *Are Big Recommendation Models Fair to Cold Users*? Available online at: https://arxiv.org/abs/2202.13607

Zhang, J., Adomavicius, G., Gupta, A., and Ketter, W. (2019). Consumption and performance: understanding longitudinal dynamics of recommender systems via an agent-based simulation framework. *Inf. Syst. Res.* 31, 76–101. doi: 10.1287/isre.2019.0876

Zhang, M., and Hurley, N. (2008). "Avoiding monotony: Improving the diversity of recommendation lists," in *Proceedings of the 2008 ACM Conference on Recommender Systems, RecSys '08* (Lausanne), 123–130.

Zhang, Y. C., Séaghdha, D. O., Quercia, D., and Jambor, T. (2012). "Auralist: Introducing serendipity into music recommendation," in *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, WSDM* '12 (Seattle), 13–22.

Zheng, Y., and Wang, D. X. (2022). A survey of recommender systems with multi-objective optimization. *Neurocomputing* 474, 141–153. doi: 10.1016/j.neucom.2021.11.041

Zhou, T., Kuscsik, Z., Liu, J.-G., Medo, M., Wakeling, J. R., and Zhang, Y.-C. (2010). Solving the apparent diversity-accuracy dilemma of recommender systems. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4511–4515. doi: 10.1073/pnas.1000488107

TYPE Review
PUBLISHED 10 May 2023
DOI 10.3389/fdata.2023.1168692

Check for updates

OPEN ACCESS

EDITED BY
Emanuel Lacić,
Know Center, Austria

REVIEWED BY
Marko Tkalcic,
University of Primorska, Slovenia
Jürgen Ziegler,
University of Duisburg-Essen, Germany
Tomislav Duricic,
Graz University of Technology, Austria

*CORRESPONDENCE
Ashmi Banerjee
✉ ashmi.banerjee@tum.de

RECEIVED 17 February 2023
ACCEPTED 18 April 2023
PUBLISHED 10 May 2023

CITATION
Banerjee A, Banik P and Wörndl W (2023) A review on individual and multistakeholder fairness in tourism recommender systems. *Front. Big Data* 6:1168692.
doi: 10.3389/fdata.2023.1168692

COPYRIGHT
© 2023 Banerjee, Banik and Wörndl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A review on individual and multistakeholder fairness in tourism recommender systems

Ashmi Banerjee*, Paromita Banik and Wolfgang Wörndl

Department of Computer Engineering, TUM School of CIT, Technical University of Munich, Munich, Germany

The growing use of Recommender Systems (RS) across various industries, including e-commerce, social media, news, travel, and tourism, has prompted researchers to examine these systems for any biases or fairness concerns. Fairness in RS is a multi-faceted concept ensuring fair outcomes for all stakeholders involved in the recommendation process, and its definition can vary based on the context and domain. This paper highlights the importance of evaluating RS from multiple stakeholders' perspectives, specifically focusing on Tourism Recommender Systems (TRS). Stakeholders in TRS are categorized based on their main fairness criteria, and the paper reviews state-of-the-art research on TRS fairness from various viewpoints. It also outlines the challenges, potential solutions, and research gaps in developing fair TRS. The paper concludes that designing fair TRS is a multi-dimensional process that requires consideration not only of the other stakeholders but also of the environmental impact and effects of overtourism and undertourism.

KEYWORDS

tourism recommender systems, travel, information retrieval, multistakeholder recommendations, fairness

1. Introduction

Recommender Systems (RS) are utilized across various domains to provide personalized access to information and help users navigate through vast amounts of content. In e-commerce, media, entertainment, and other industries, they improve the user experience, increase engagement, boost sales, and drive revenue. By enhancing the discoverability of relevant items, RS ultimately lead to greater satisfaction and loyalty among users (Ricci et al., 2020).

In the past, evaluating the effectiveness of recommender systems was mainly based on their ability to cater to the needs and preferences of end users. This approach makes sense as users would not use the systems if it does not meet their interests. However, RS have seen a tremendous gain in popularity and have now impact beyond the users they were initially designed for. Therefore, it is important to note that in many cases, the end user is not the only stakeholder impacted by the recommendations. Other users, product providers, and the system's goals should also be taken into account. This has led to the inclusion of objectives like fairness and balance in the evaluation process, even if they may not align with individual preferences. Focusing solely on the end user limits the ability to consider the concerns of the other stakeholders in the design and algorithm of recommender systems (Abdollahpouri et al., 2020). Hence, a fair recommender system is evaluated from various stakeholders' perspectives, making it a complex and multi-faceted process (Burke, 2017).

The widespread adoption of RS in the travel industry has made trip planning easier for travelers by offering personalized recommendations for destinations, accommodations, activities, etc. (Isinkaye et al., 2015). Tourism Recommender Systems (TRS) stand out from other RS domains due to their susceptibility to dynamic factors that are subject to frequent changes. For instance, changes in seasonality or travel regulations can have a significant impact on travel plans (Balakrishnan and Wörndl, 2021). Furthermore, it also involves capacity-limited items, including airline seats, hotel rooms, and tickets to events, further aggravating the complexity of the domain (Abdollahpouri and Burke, 2021).

In the realm of tourism, where recommendations can greatly impact not only the end user but also the local community and the environment, it becomes even more crucial to evaluate recommender systems from multiple perspectives and strive for fairness in their recommendations. The travel and tourism domain is complex, encompassing various stakeholders beyond just the traveler, such as transportation providers, host destinations, and information platforms, each with their own needs and goals (Abdollahpouri et al., 2020). Additionally, while constructing a fair TRS, it is important to take into account the environmental impact of tourism. Tourism and the environment are intertwined in a complex relationship that includes activities that can have both negative and positive impacts. On one hand, tourism can contribute to environmental protection and conservation, raise awareness of environmental values, and provide funding for natural areas. On the other hand, it can also have adverse effects such as contributing to climate change, depleting natural resources, causing overtourism or undertourism, etc. (Camarda and Grassini, 2003; Gössling, 2017).

A well-designed TRS can be particularly beneficial in controlling the influx of tourists to a region. Such control is essential in addressing two related problems that have become increasingly prevalent in recent years: overtourism and undertourism. The growth of low-cost aviation, cheap transportation, social media popularity, and home-sharing platforms like Airbnb[1] have led to a surge in visitors to popular destinations, resulting in overtourism. At the same time, there are under-explored destinations that suffer from undertourism due to a lack of infrastructure, publicity, and accessibility (Gowreesunkar and Vo Thanh, 2020). Both over and undertourism have several negative consequences. Overtourism endangers the preservation of the city's historic center and has negative consequences for the environment, residents, and tourists' experiences, making it challenging to find reasonably priced housing in these cities (Dastgerdi and De Luca, 2023). Cities in Europe such as Venice, Barcelona, Rome, and Dubrovnik are grappling with the effects of overtourism (Dodds and Butler, 2019). A lack of tourists on the other hand can have adverse effects as well as experienced during the Covid-19 pandemic outbreak. The pandemic had a profound impact on the tourism industry, causing severe disruptions to the tourism and hotel industries (Hao et al., 2020; Galí Espelt, 2022).

1 https://www.airbnb.com/

To help mitigate these and other problems TRS should be designed to take into account the interests of all stakeholders, advocate for sustainable tourism practices, and encourage responsible tourism while providing recommendations to users. To this end, our work makes the following three contributions:

- We highlight the main fairness criteria and categorize stakeholders based on the ones that apply to them.
- We review state-of-the-art research on TRS fairness from multiple stakeholder perspectives.
- Finally, we outline the challenges, potential solutions, and research gaps to lay the foundation for future research in developing fair TRS.

The paper is structured as follows: we begin with an overview of fairness in RS and the stakeholders involved in TRS in Section 2.1 and Section 2.2. Next, in Section 2.3, we explain our methodology for identifying relevant papers for our survey and provide some statistical information on the papers reviewed. We then delve into the concept of individual or intra-stakeholder fairness in Section 3 and examine works that focus on multiple stakeholders simultaneously in Section 4. Finally, in Section 5, we conclude the paper by discussing the challenges encountered by TRS and potential solutions to address them.

2. Terminology

2.1. Fairness in RS

In an era, where data drives decisions, it is crucial to examine if algorithms may discriminate based on gender, ethnicity, or other protected attributes. Multiple studies have investigated fairness in decision-making systems based on Machine Learning methods (Pedreshi et al., 2008; Zemel et al., 2013; Hardt et al., 2016; Zafar et al., 2017; Speicher et al., 2018), and Information Retrieval (Castillo et al., 2017; Yang and Stoyanovich, 2017; Biega et al., 2018; Celis et al., 2018; Singh and Joachims, 2018).

A multitude of fairness notions has been studied to ensure that algorithmic decisions are fair. They can be divided into *individual* and *group fairness* notions. Group fairness ensures fair treatment of similar subjects within the different groups based on protected attributes such as race or gender (Masthoff and Delić, 2012). Individual fairness assesses whether individuals are treated fairly by ensuring that similar subjects receive similar decision outcomes (Dwork et al., 2012).

The concept of fairness applies to RS too. RS offer personalized access to a vast amount of content across domains like e-commerce, social media, news, travel, and more, finding relevant information and avoiding information overload (Abdollahpouri et al., 2020). They are usually evaluated for recommendation accuracy, i.e., their ability to provide a list of items that meet the user's needs. However, increased awareness of fairness and bias issues in algorithmic decision-making (Romei and Ruggieri, 2014) have led researchers to focus on fairness aspects in RS evaluations (Kamishima et al., 2013; Burke, 2017; Serbos et al., 2017; Xiao et al., 2017; Yao and Huang, 2017; Burke et al., 2018; Liu and Burke, 2018; Steck, 2018; Abdollahpouri et al., 2020).

While the notions of individual and group fairness can be applied to RS as well, fair machine learning differs from fairness in RS through the multi-sided nature of the latter. Fairness in recommendation systems is often a multi-sided concept that takes into account the needs and perspectives of multiple stakeholders (Burke, 2017). In other words, there may be multiple fairness-related criteria at play in determining fair outcomes and these outcomes cannot be evaluated based solely on the results for one side of a transaction. In RS, a *stakeholder* is any group or individual that can be affected by or can affect the delivery of recommendations (Abdollahpouri et al., 2020). Therefore, a *multistakeholder* RS should serve the goals of all stakeholders involved. However, in practice, this is often not the case, which is attributed to the existence of different biases in RS.

RS can exhibit the following three types of common biases: *popularity*, *exposure*, *ranking*, or *position bias*. *Popularity bias* is a major fairness concern in recommendation systems. It refers to the tendency of the system to recommend items that are often popular among users, regardless of the individual preferences of a particular user (Abdollahpouri et al., 2019a). This can often result in less popular items being disfavored, leading to unfair recommendations in terms of the exposure given to different items of varying popularity, known as *exposure bias* (Abdollahpouri and Mansoury, 2020). In recommender systems, rankings of items play an integral role in the decision-making process. As ranking positions influence the amount of attention received by the ranked items, biases in rankings can lead to the unfair distribution of resources and opportunities (Biega et al., 2018). This type of bias is known as the *ranking bias* or *position bias* in the literature.

As pointed out by Buet-Golfouse and Utyagulov (2022), fairness definitions often vary based on domains and context. To study fairness in multistakeholder recommender systems, it is important to identify the stakeholders who should receive fair treatment, quantify any harms that may occur, and analyze metrics for measuring and minimizing these harms (Ekstrand et al., 2020). This process of defining an objective function involves taking a concern (in this case, reducing representational harm) and translating it into a specific framework and metric (Ekstrand et al., 2020). The resulting metric should also measure the usefulness i.e. the utility of the recommendations for the user. In our work, we conceptualize the *utility of a recommendation result for a multistakeholder system as its usefulness for each stakeholder*.

However, it is important to note that this process of defining a metric is inherently limited and may result in trade-offs. These limitations and trade-offs do not necessarily render the fairness construct invalid. All fairness constructs come with their limitations and trade-offs and there is no universally accepted definition of fairness (Narayanan, 2018; Ekstrand et al., 2020). In our work, *a multistakeholder recommender system is considered to be fair if it minimizes any bias or circumstance that may result in disfavored outcomes for each stakeholder.* This implies that a fair multistakeholder RS may have to consider trade-offs in the respective stakeholder concerns.

2.2. Stakeholders in TRS

In the tourism industry, the traveler is not the only stakeholder involved. Every service that is part of the traveler's journey, including transportation providers, host destinations, and information platforms, also has a stake in the industry (Abdollahpouri et al., 2020). Optimizing recommendations for the consumers' experience can often align with and benefit the goals of the providers, such as increased sales or higher engagement. However, there may also be situations where achieving the goals of one stakeholder may come at the expense of another stakeholder's goals, creating potential trade-offs (Jannach and Bauer, 2020).

Following the classification of common stakeholders encountered in a generic multistakeholder recommender system we generalize the type of stakeholders encountered in common touristic recommendation scenarios into the following classes. Our categorization is inspired by the work of Balakrishnan and Wörndl (2021).

- *Consumers*: the end users who receive or want to receive recommendations to plan their trips, such as tourists, business travelers, airline passengers, etc.
- *Item Providers*: the entities that provide the consumers with the recommended facility for their trips, such as hotels, resorts, rentals, amusement parks, airlines, tour operators, and vacation companies.
- *Platform*: the recommender system itself, such as flight booking platforms, vacation recommenders, city information systems, travel sites, e-commerce sites, hotel platforms, and similar systems.
- *Society*: it represents the environment and entities or groups that are affected by the tourism activity but are not directly part of the TRS. This can include the local environment, city authorities, municipal councils, local businesses, and Destination Management Organizations (DMOs).

Although the aforementioned stakeholder categorization seems plausible, stakeholder relationships, in reality, can be more complex. For instance, in the context of tourism's value chain, the providers of final services (e.g., hotels), travel companies, online travel platforms, and even travel agencies can be further subdivided despite being grouped as item providers. This grouping may create a false impression that they all share the same interests, which is not the case and could impact fairness for these different groups. However, this is the most logical and simplified approach to structuring the stakeholders based on earlier works by Abdollahpouri et al. (2020), Jannach and Bauer (2020). On the other hand, the inclusion of society as a stakeholder in tourism by Jannach and Bauer (2020); Balakrishnan and Wörndl (2021) is a novel and appropriate perspective. This viewpoint adds an interesting dimension to the functioning of multistakeholder tourism, emphasizing the crucial issue of reducing the environmental impact caused by tourist activities.

To understand the stakeholder interplay, let us consider the example of a hotel booking scenario on a platform like

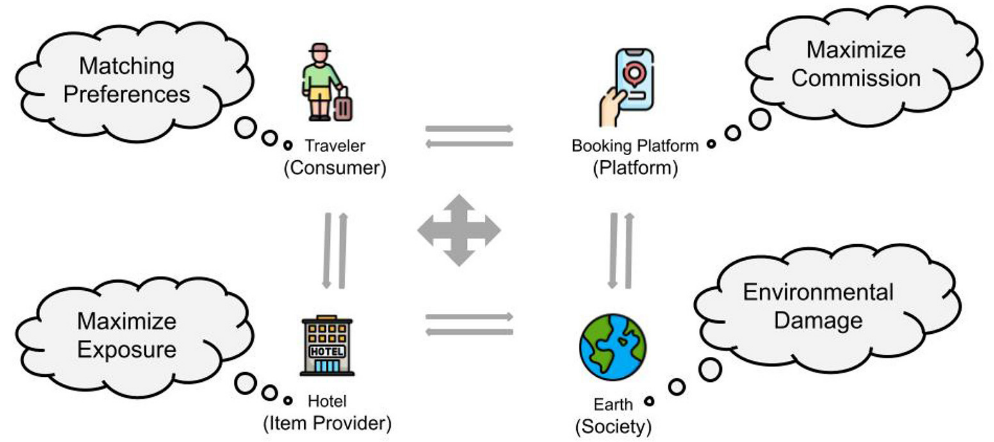

FIGURE 1
The multistakeholder environment in a hotel booking scenario adapted from Abdollahpouri and Burke (2019).

Booking.com[2] in Figure 1. Here, we can observe all four major stakeholders as shown: (1) end users or travelers who are searching for accommodation in the city during the specified period, (2) the hotels that are being recommended, (3) the booking platform itself that is providing the recommendations for hotels and 4) Society i.e. city authorities, municipal councils, and DMOs who must ensure that the city is not over-crowded and the environment is not compromised.

Travelers want to find hotels that match their preferences, hotels want fair exposure to attract guests, and booking platforms want to maximize the commission received from the hotels and maintain long-term relationships with both users and hotel providers. All stakeholders are dependent on each other for their economic well-being, and therefore the booking platform must take all stakeholders' preferences into account when generating recommendations. Additionally, society plays a key role in ensuring minimal environmental impact and avoiding overcrowding in the city. As a responsible platform, Booking.com should take into consideration the concerns of indirectly affected actors from society. This example reinstates the domain of tourism as a prime use case for studying multistakeholder recommender systems, where different stakeholders interact with one another directly or indirectly. As the example shows, these stakeholders are often interdependent for their existence. Furthermore, in certain situations, stakeholders may play multiple roles and should be considered separately as distinct entities, as discussed by Balakrishnan and Wörndl (2021).

We use analogous terminology as Abdollahpouri and Burke (2019) to demonstrate the close connection between multistakeholder recommendation and multi-sided fairness. We categorize fairness into four groups— *C-Fairness*, which focuses on consumers and encompasses *individual* and *group discrimination*; *I-Fairness*, which targets item providers and deals with *popularity bias* and *exposure bias*; *P-Fairness*, which concentrates on platforms and addresses *ranking bias*; and *S-Fairness*, which takes into account the impact on society through *sustainability*. These groups provide an intra-stakeholder perspective on fairness along with their respective key fairness criteria. Each group has been further studied with different fairness attributes as summarized in Table 1. The overlapping stakeholder scenarios have been addressed in Section 4.

2.3. Research methodology

The following section outlines our approach to identifying pertinent papers for the survey. We will then provide a brief overview of how our survey builds upon previous research in this field.

Firstly, we developed a methodology to identify relevant papers for our survey. We began by using predefined search terms and explicit inclusion and exclusion criteria to query the Google Scholar web search engine. Additionally, we employed a snowballing technique and relied on researcher experience to identify any relevant papers that were not captured by our initial search.

To ensure that we covered a broad range of works in an emerging field with inconsistent terminology, we used the following keywords: *tourism*, *fairness*, *multistakeholder*, and *recommender systems*. Later, to identify more studies specific to the tourism industry, we expanded our search terms to include relevant terms to tourism platforms such as *Airbnb*, *TripAdvisor*, *Yelp*, and Booking.com.

Finally, we manually reviewed the resulting papers to determine if they met the following criteria for inclusion in our survey:

- It had to include at least one fairness criterion or bias in RS as identified in Table 1.
- It has to be within tourism or a comparable domain.
- It has to be about RS, ranking, or information retrieval.
- It has to be published in the last decade except for two papers that were included due to their significant conceptual contributions to the field of ranking, rather than their specific use cases.

2 https://www.booking.com/

TABLE 1 Table summarizing related works in fairness from different stakeholder perspectives, that directly or indirectly contribute to Tourism Recommender Systems.

Fairness type	Stakeholder focus	Main fairness criteria	References
C-Fairness	Consumers/ End-Users	Individual Discrimination	Edelman et al. (2017); Serbos et al. (2017)
			Herzog and Wörndl (2019); Jaeger and Sleegers (2020)
			Zhang et al. (2022)
		Group Discrimination	Delic et al. (2018); Mansoury et al. (2019)
			Rahmani et al. (2022a)
I-Fairness	Item-providers/ Producers	Popularity Bias	Jannach et al. (2015); Fu et al. (2021)
			Pala (2021); Wei et al. (2021)
			Lin et al. (2022); Tacli et al. (2022)
			Zhou et al. (2020); Yalcin and Bilge (2022)
		Exposure Bias	Abdollahpouri and Mansoury (2020); Banerjee et al. (2020)
			Khenissi and Nasraoui (2020); Gupta et al. (2021b)
			Yang et al. (2021)
P-Fairness	Platforms/ Systems	Ranking Bias	Biega et al. (2018); Grbovic and Cheng (2018); Lahoti et al. (2019)
			Gunawardena and Sarathchandra (2020); Kokkodis and Lappas (2020)
			Li (2020); Mavridis et al. (2020)
			Zhu et al. (2020); Gupta et al. (2021a)
			Kangas et al. (2021); Gupta et al. (2022)
S-Fairness	Society	Sustainability	Patro et al. (2020b); Pachot et al. (2021)
			Merinov et al. (2022)

In this process, we identified a total of 66 papers, which we divided into TRS and non-TRS categories based on the domain in which they focussed. Figure 2A illustrates the number of papers on fairness in TRS published per year and from the visualization, it is evident that fairness in TRS is a relatively nascent field, with the most recent research dating back to 2014.

The resulting papers for TRS were systematically analyzed based on four distinct aspects: the specific fairness criteria or bias being addressed, solutions proposed, results evaluated, and datasets analyzed. According to our analysis presented in Figure 2B, previous research has primarily focused on fairness to consumers, item providers, and platforms, with very little attention given to fairness to society as a stakeholder. We further discuss this in detail in Section 3. As depicted in Figure 2C, various datasets from the travel and tourism domain were utilized in the studies analyzed.

It is worth noting that fairness in RS has been extensively surveyed in literature in recent years, as shown by authors such as Deldjoo et al. (2022) and Wang et al. (2022), who provide in-depth reviews of related concepts and work on the particular topic. However, our study aimed to provide a comprehensive understanding of fairness from multiple stakeholders' perspectives within TRS, with a specific focus on society. As a result, our survey differs from previous surveys in that our objective was to investigate and identify research gaps in the current state of existing research for TRS.

3. Individual stakeholder fairness in TRS

In this section, we discuss related papers including fairness concerns that were gathered according to the explained methodology in Section 2.3. Each subsection focuses on stakeholder fairness with respect to the primary fairness criteria presented in Table 1, and it examines relevant literature falling under those criteria. The approaches outlined in Table 1 primarily address one primary stakeholder and fairness criterion, but they can be applied, to some extent, to other stakeholders as well. For instance, while popularity bias and exposure bias can have an impact on the platform itself, their primary effects are on the providers of the items. Hence, for the sake of simplicity, we have chosen to address the primary fairness criteria to a single stakeholder.

3.1. Consumer fairness: individual and group discriminations

Consumer Fairness (C-Fairness) refers to the need for a recommender system to consider the different effects of its recommendations on the protected or sensitive attributes of its users, such as age, gender, and nationality. It also encompasses any fairness concerns the system may have concerning its users (Sonboli et al., 2021).

C-Fairness can appear on an individual as well as a group level. Individual Fairness refers to treating similar individuals in a similar way (Dwork et al., 2012). In a group recommender system, this means considering the preferences of all group members fairly and not ignoring any individual's preferences (Masthoff and Delić, 2012).

Despite a significant amount of research on C-Fairness in other application domains, such as music recommendations (Dinnissen and Bauer, 2022), the field of travel and tourism has seen relatively little analysis of this topic. In the tourism industry, researchers often examined the effect of a user's gender and the business category on various outcomes (Mansoury et al., 2019). For example, a study on Airbnb revealed that hosts who had never hosted African American guests were less likely to accept guests with African American names compared to those with White names (Edelman et al., 2017). Another analysis showed that non-White hosts charge 2.5–3% lower prices for similar listings, while

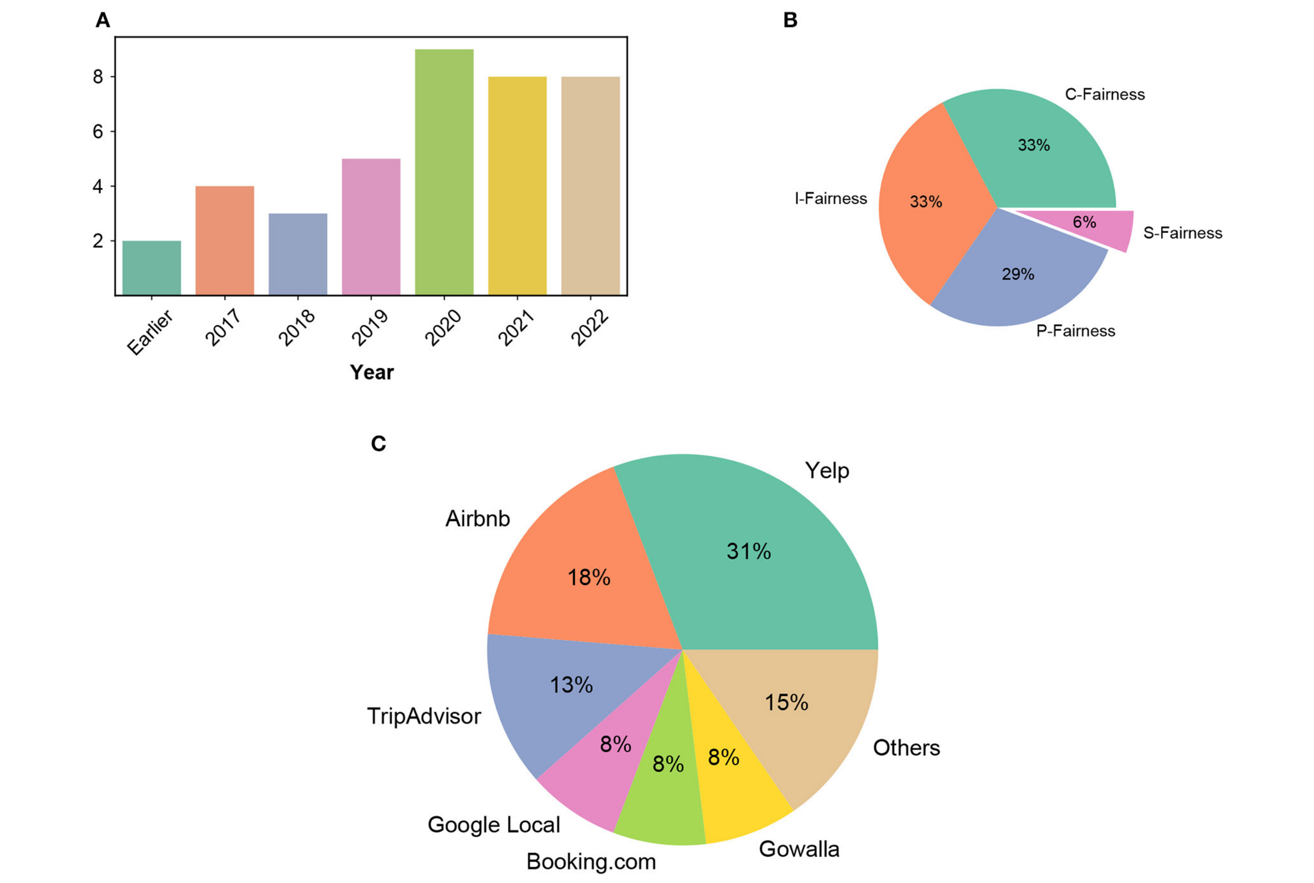

FIGURE 2
(A) The number of papers on fairness in TRS published per year. The total number of papers amounts to 66. **(B)** Percentage of papers with each stakeholder focus in TRS. **(C)** Percentage of papers analyzing each dataset in TRS.

Black and Asian hosts charge approximately 5–7% and 4–6% less respectively (Jaeger and Sleegers, 2020). It was also found that discrimination between hosts and guests on Airbnb is reciprocal, with specific topics in reviews and self-descriptions significantly associated with discrimination (Zhang et al., 2022).

Group recommender processes in the tourism industry have been explored by Delic et al. (2018). The study aimed to observe the evolution of user preferences and interactions as a group during a tourism decision-making task. The authors also provided a comprehensive description of the study's data collection procedure, which can be utilized for further analysis to gain a deeper understanding of group decision-making processes.

Rahmani et al. (2022a) explored the impact of adding contextual information (such as geographic, temporal, social, and categorical details) on the quality of point-of-interest recommendations. They focused on four aspects: accuracy, novelty, diversity, coverage, fairness, and interpretability. The authors developed a linear regression approach for combining contextual information from different sources and applied it to two datasets (Gowalla[3] and Yelp Challenge[4]) to assess the fairness of recommendations for both active and inactive users and popular and less popular items. Their results suggest that context-aware recommendation methods tend to be fairer to both users and item providers compared to traditional collaborative filtering methods.

While most of the aforementioned studies aim at being fair to a group of users, the work by Serbos et al. (2017) propose envy-free tour package recommendations for travel booking sites to ensure each individual is satisfied with their recommendation, demonstrating their findings on the Yelp Challenge dataset. An analogous concept was covered by Herzog and Wörndl (2019), where they focused on individual fairness in user groups and addressed the recommendations of points of interest (POIs) based on group preferences often tend to be unfair for some group members. The authors proposed a distributed Group Recommender System (GRS) that aggregates all group members' individual preferences fairly with the option to share one display for all members to openly discuss their preferences. Their study results showed that the approach could deliver fairer recommendations to groups with close relationships between members as they felt more comfortable specifying travel preferences as a group. Whereas groups with looser connections preferred to use separate devices to specify their preferences individually and to leave the preference aggregation to the GRS.

3 http://snap.stanford.edu/data/loc-gowalla.html

4 https://www.yelp.com/dataset

3.2. Item-provider fairness: popularity bias and exposure bias

The item providers are the entities that offer or support the recommended items. A recommender system that has an item provider fairness (I-Fairness) requirement should treat these providers of items in an equitable manner (Abdollahpouri and Burke, 2019; Abdollahpouri et al., 2020). Ensuring I-Fairness is particularly important in multistakeholder systems, as not recommending an item of quality can lead to economic hardship for the item provider and can also negatively impact market diversity by allowing certain providers to dominate (Banerjee et al., 2020). This section centers on the unfair treatment of item providers resulting from popularity bias and exposure bias.

In the context of recommender systems, an item's likelihood of being recommended to a user is not only based on the user's preferences but also on the item's popularity and visibility on the platform. Popularity bias is a common data bias that affects recommender systems, causing them to favor more popular items over less popular ones (Bellogín et al., 2017). This can lead to a lack of representation and fairness for less popular items or items that are only popular among small groups of users (Park and Tuzhilin, 2008). This bias can also be seen as unjust to the providers of less popular or new items as few users rate them (Abdollahpouri et al., 2019a). Furthermore, a market that is dominated by popularity bias will not allow room for the exploration of new and obscure products and will be limited by a small number of well-known item providers leading to a lack of diversity, stifling innovation and creativity, ultimately limiting the market (Abdollahpouri et al., 2019a).

Various provider-side bias mitigation strategies have been suggested by other researchers. These include statistical parity (Yang and Stoyanovich, 2017), balanced neighborhoods (Burke et al., 2018), statistical independence (Kamishima et al., 2018), pairwise comparison (Beutel et al., 2019), data re-sampling (Ekstrand et al., 2018; Rastegarpanah et al., 2019; Boratto et al., 2021), etc. The study of popularity biases from the item providers' perspective remains a widely researched topic not only in tourism but also in other domains (Kamishima et al., 2014; Abdollahpouri et al., 2017; Abdollahpouri, 2019).

Lately, there has been a growing focus on analyzing popularity biases in TRS. The study by Jannach et al. (2015) using a hotel proprietary dataset from HRS.com[5] found that popular recommendation techniques prioritize a small portion of items or top sellers, and have limited accuracy. The popularity bias in Yelp data was analyzed by Zhou et al. (2020). They concluded that models relying solely on positive reactions such as purchases or clicks result in less personalized recommendations and heightened popularity bias. To overcome this, they suggest incorporating implicit feedback and user-generated reviews, which provide a wealth of preference information for each user. The use of user-generated reviews was also explored by Pala (2021) using TripAdvisor data to compare top-ranked and least-ranked hotels. They found little difference in online review sentiment for both types of hotels, indicating that popularity does not solely reflect quality (Ciampaglia et al., 2018). The cause-effect relationship of popularity bias was addressed by Wei et al. (2021), where they estimated the direct effect of item properties on the ranking score, and then removed it to eliminate popularity bias. Their strategy was proven effective through extensive experiments on multiple real-world recommendation datasets, including Yelp and Gowalla.

Debiasing frameworks for addressing popularity bias in Conversational Recommender Systems (CRS) have been proposed by Fu et al. (2021) and Lin et al. (2022). through various debiasing frameworks. The former introduced metrics for quantifying popularity bias in CRSs, along with a debiasing framework, while the latter presented a framework that balances recommendation performance and item popularity in the CRS environment by combining dialogue context and historical user information. Their experiments on the Yelp dataset demonstrated a successful balance between the effectiveness of recommendations and the popularity of the items in the conversational recommendation system setting.

Popularity bias can often result in less popular items being disfavored, leading to unfair recommendations in terms of the exposure given to different items of varying popularity, known as exposure bias (Abdollahpouri and Mansoury, 2020). They propose metrics to quantify exposure bias from the perspective of the users and the providers by evaluating their research on Last. FM[6] and MovieLens[7] datasets. They show that when the recommendations are calibrated for the users in terms of popularity it will also benefit the providers by providing them with the exposure that they deserve, further reinforcing the idea that RS should be evaluated from multistakeholder viewpoints. Even though their work is based on the Last.Fm and MovieLens data, it can be translated into the travel domain such as for destinations/POIs recommendations displayed on different platforms. The studies by Tacli et al. (2022) and Yalcin and Bilge (2022) similarly address popularity bias in Yelp data by analyzing users' preferences for popular items. Tacli et al. (2022) suggest evaluating users' actual tendencies toward item popularity to provide more accurate individual recommendations.

The work by Banerjee et al. (2020) quantify exposure bias arising from popularity and position bias in the case of location-based searches. Their experimental evaluation of multiple real-world datasets from Google, Yelp, and Booking.com reveal the existence of exposure disparity on these platforms. Exposure bias has been addressed from a causality perspective by Yang et al. (2021). They argue that a combination of deep learning techniques along with causal inference is an effective method to mitigate exposure bias in RS. The studies by Khenissi and Nasraoui (2020) and Gupta et al. (2021b) also propose novel methodologies to model and mitigate exposure bias. Even though their work is demonstrated on the MovieLens dataset and for citation link recommendations respectively, the concept and methodology can be translated to the domain of tourism as potential strategies to minimize exposure bias in TRS.

Gunawardena and Sarathchandra (2020) suggest the use of deep neural networks to create a digital menu and personalize food item recommendations for customers, allowing them to make

5 https://www.hrs.com/

6 https://www.last.fm/

7 https://grouplens.org/datasets/movielens/

informed decisions. While the research does not specifically address fairness in recommendations, the approach could potentially be applied to the tourism industry as a means of providing fair food recommendations.

3.3. Platform fairness: ranking bias

Online platforms greatly impact offline experiences, such as selecting a tourist destination (Huang et al., 2018). The visibility of items on the platform is crucial to their success (Abdollahpouri and Mansoury, 2020). Items at the top of search results attract more attention, while those lower down may miss out on business opportunities (Craswell et al., 2008; Ursu, 2018). Additionally, platforms may be tempted to favor certain items more due to the commissions they receive from the item providers (Jannach and Bauer, 2020), which can lead to an unfair distribution of items on the platform and negatively impact some of its stakeholders. It's important for platforms to ensure fair item ranking to promote diversity in recommendations and ensure platform fairness. Unfair ranking can negatively impact stakeholders and erode trust in the platform. This paper analyzes the impact of unfair ranking on P-Fairness, focusing on platforms and item providers as individual stakeholders.

Apart from studies that have examined the impact of search rankings and position bias in different information retrieval scenarios [such as (Fortunato et al., 2006; Craswell et al., 2008; Chuklin et al., 2015; Baeza-Yates, 2018; Ursu, 2018; Geyik et al., 2019; Draws et al., 2021)], there have also been studies that have aimed to develop a fair ranking strategy specifically for the travel and tourism industry. TripAdvisor[8], Airbnb, Booking.com, and Yelp are among the travel platforms that have been studied concerning the concept of fair ranking.

The authors of Li (2020) studied TripAdvisor data and found that Learning-to-Rank models based solely on implicit user feedback (such as clicks) can lead to bias. They proposed a method that takes into account the user's evaluation of all hotels above the clicked result and samples hotels below it based on their propensities. Their online experiment on TripAdvisor showed significant improvement in the search ranking using this method. Grbovic and Cheng (2018) propose search ranking methods tailored to Airbnb, using embedding techniques to personalize recommendations in real-time and effectively suggest home listings. Biega et al. (2018) introduced a notion of amortized fairness in ranking, which accumulates fairness over multiple rankings, resulting in improved individual fairness with high-ranking quality according to their study on Airbnb data. Gupta et al. (2021a) suggested re-ranking methods for online post-processing based on ranked batches of items, balancing fairness and utility, and performing well on Airbnb data. Lastly, Lahoti et al. (2019), focus on reconciling the fairness and utility of Airbnb data and propose a framework that results in individually fair learning-to-rank results. Mavridis et al. (2020) shed light on the multiple factors beyond the choice of algorithm that must be addressed for creating a machine-learned ranker in a large-scale commercial setting such as Booking.com. The authors suggest that their research could serve as guidance for applying machine learning to ranking problems. Another study by Kangas et al. (2021) address fair ranking in TRS platforms from a user experience perspective by developing a framework in Booking.com. This framework allows for the dynamic addition and removal of items, ensuring that new items have a fair chance, and enables recommendation blocks to be ranked in the most relevant order for the user interface. In their study, Zhu et al. (2020) propose a debiased ranking model that uses statistical parity and equal opportunity to mitigate item under-recommendation bias in personalized ranking systems. Their experiments on three publicly available datasets, including Yelp, demonstrate significant bias reduction compared to current state-of-the-art methods.

The concept of P-Fairness has also been explored for restaurant recommendations. For example, Kokkodis and Lappas (2020) proposed a fair ranking system for online platforms by examining the impact of the popularity-difference bias on online restaurant reviews. This bias stems from the difference in popularity between the reviewer's hometown and the destination being reviewed, which can lead to conflicting opinions on the effect of this bias on assigned ratings and review sentiment. The authors' analysis of a large set of restaurant reviews from a major online platform reveals a significant impact of this bias on restaurant ratings. They suggest that recognizing this bias can help online platforms improve their ranking systems, resulting in improved satisfaction for reviewers and more diverse recommendations for top restaurants.

Moreover, Gupta et al. (2022) present a novel solution to ensure fairness in food delivery services through the FairFoody algorithm. This algorithm uses delivery data to allocate fair income distribution among agents, while also ensuring timely deliveries. FairFoody's approach is unique in its focus on fairness in income distribution among agents, rather than just the recommendations. This could have potential applications in the tourism industry, such as fair allocation of resources among food vendors at a tourist destination.

To summarize, fair ranking in online platforms is essential in promoting diversity and building trust among customers and item providers. The online platforms have an ethical and moral responsibility to ensure that their recommendations are fair to all stakeholders. This fairness should be evident not only in terms of visibility and exposure but also in the ranking process to promote fair competition. Moreover, while there is a significant body of research focused on fair rankings in the context of hotel and restaurant recommendation platforms, there are few studies that address this concern for other tourism-related issues, such as trip planning or route optimization. Therefore, exploring fair ranking or ensuring P-Fairness in these areas presents a promising avenue for future research.

3.4. Societal fairness: sustainability

The impact of tourism extends beyond active participants to affect the local environment and businesses. Therefore, constructing a fair TRS requires considering sustainable recommendations. World Tourism Organization and United

8 https://www.tripadvisor.com/

Nations Development Programme define *sustainable tourism* as "*tourism that takes full account of its current and future economic, social and environmental impacts, addressing the needs of visitors, the industry, the environment, and host communities*" (Gössling, 2017). Societal Fairness, also known as S-Fairness, focuses on meeting the needs of non-participating stakeholders in tourism, such as residents who may be affected by issues such as housing prices and congestion. In this context, we use the terms S-Fairness and Sustainability interchangeably.

Achieving sustainability in tourism requires various types of interventions, including municipal policies and regulations. To ensure sustainability in TRS, possible interventions include reducing the environmental impact of tourism, balancing the tourist load, promoting public transportation, encouraging carpooling, and supporting sustainable business practices. However, the idea of generating sustainable recommendations is a relatively new concept with limited literature available. Current literature on TRS focuses on regulating the number of tourists traveling to a destination to control the impact of tourism, particularly in preventing phenomena like over and undertourism.

The terms over and undertourism are used to describe situations where a destination is overwhelmed by too many tourists or lack tourists, respectively. Overtourism has become increasingly prevalent due to factors such as affordable transportation, home-sharing services, and exposure disparity caused by social media/recommendation technologies, leading to negative impacts on the environment, residents, and tourists' experiences (Camarda and Grassini, 2003; Rabanser and Ricci, 2005; Hospers, 2019; Dastgerdi and De Luca, 2023). Undertourism, on the other hand, occurs in lesser-known destinations with insufficient infrastructure, publicity, and accessibility, resulting in economic disadvantages (Gowreesunkar and Vo Thanh, 2020).

The idea of developing sustainability-driven recommender systems has recently received attention in the literature. For example, Merinov et al. (2022) have explained how recommender systems can potentially be used as a medium to introduce under-visited areas and strategically control tourists in over-visited areas through a case study on an Italian village. The authors proposed a multistakeholder utility model for travel itinerary optimization that protects popular destinations from overpopulating and less mature destinations from under-populating by distributing tourists throughout different points of interest (POIs) while preserving user satisfaction. The model used user preferences from the consumer side and time and occupancy of POIs from the environment side as two objectives and optimized the trade-off between the two using a greedy breadth-first search graph method to recommend optimal itinerary routes to users.

While research on the topic of over and undertourism in TRS is limited, the COVID-19 pandemic has sparked interest in utilizing RS to address these challenges and promote sustainable production systems. The pandemic has presented multiple challenges for businesses, including the need to maintain social distancing in public spaces such as restaurants and other venues. This has resulted in overcrowding in some places, compromising customer safety, and very low footfall in others, jeopardizing their economic sustainability. Patro et al. (2020b) addressed this issue by formulating it as a multi-objective optimization problem and mapping it to a bipartite matching with a polynomial time solution. Their experiments on real-world datasets from Yelp and Google Local[9] have demonstrated that their model improves business sustainability, safety, and utility goals.

In addition, the pandemic has drawn attention to the importance of sustainable production methods in local businesses, with a focus on prioritizing the rights of local communities over the desires of tourists and the profits of tourism companies (Higgins-Desbiolles et al., 2019). To address this, Pachot et al. (2021) have proposed a novel recommender system for companies that takes into account territorial policies, while promoting diversity and providing a competitive advantage for providers. The objective of this system is not only to promote business growth, but also to consider factors such as economic growth, productive resilience, securing necessities, and sustainable production for local authorities. This approach offers a fresh perspective on the evaluation of S-Fairness in recommendations by emphasizing the involvement of local authorities (society), providing insights into an area of fairness in recommendations that have previously been unexplored.

To summarize TRS may have unintended consequences for other stakeholders who are indirectly involved in the process of recommendation. This highlights the importance of a holistic recommendation process that considers the perspectives and interests of all parties, including society. Initial research has indicated that TRS has the potential to effectively manage the allocation of limited resources, but its potential for addressing tourism-related concerns remains an open question.

4. Multistakeholder fairness in TRS

In certain applications, multiple fairness concerns may arise simultaneously for different stakeholders. Thus, a system may have any combination of the previously mentioned fairness considerations at play at once, such as for both consumers and providers, but also any other combination of stakeholders. Moreover, often the stakeholder concerns are conflicting, making it difficult to satisfy the specific concern of a single stakeholder. For example, a rental platform such as Airbnb and its rentals (item providers) share a common objective of avoiding position or popularity bias. To optimize the ranking of the rentals, it's necessary to simultaneously consider both P-Fairness and I-Fairness in this case. Therefore, in this section, we review methods that have simultaneously addressed more than one stakeholder in their fairness criteria.

To resolve the challenge of ensuring fairness toward multiple stakeholders, many studies adopt a multi-criteria optimization approach. This method involves optimizing a utility function that accounts for multiple criteria and preferences of various stakeholders while aiming to maintain a minimal trade-off in personalization. This approach is commonly used in other domains such as movies or

9 https://developers.google.com/maps/documentation/places/web-service/overview

music (Bouveret et al., 2016; Burke et al., 2016; Liu et al., 2019; Sühr et al., 2019; Patro et al., 2020a; Ranjbar Kermany et al., 2021), but has not yet been widely adopted in the field of tourism.

We have grouped the literature into three categories in Table 2: (1) works that specifically deal with fairness in TRS from multiple stakeholder perspectives, (2) works within the TRS domain that address multi-criteria recommendations, and (3) recent studies in other domains that address fairness in a multistakeholder scenario and can be adapted to the tourism industry.

4.1. Fairness in TRS

Fairness in TRS should be addressed from a multi-sided perspective owing to the involvement of multiple stakeholders in the system. In Section 3, the primary focus was on addressing fairness concerns for a single stakeholder. In this section, the focus shifts toward simultaneously optimizing fairness concerns for multiple stakeholders. The studies reviewed in this section use multi-objective optimization frameworks to generate fair recommendations in the tourism domain.

In the context of location-based recommendations Rahmani et al. (2022b) focus on addressing user fairness and item fairness for point of interest (POI) recommendations. They classify users into advantaged and disadvantaged levels based on their activity level and divide items into short-head, mid-tail, and long-tail groups to study their exposure in the recommendation list for users. They examine the interactions between different factors such as the unfairness of users (C-Fairness), the unfairness of popular items (I-Fairness), and the personalization offered by the recommender system (P-Fairness). Through evaluating their algorithms on publicly available datasets from Yelp and Gowalla, they found that most well-performing models suffer from popularity bias (provider unfairness). Furthermore, their study highlights that most recommendation models are unable to simultaneously satisfy both consumer and producer fairness, indicating a trade-off between these variables possibly due to natural data biases. Weydemann et al. (2019) explore the quantification of fairness in location recommendations. Their study focuses on different fairness aspects, and the results are based on data from Travel Data Solution, an Austrian company that equips rooms of certain hotels in Austria with cellular-based mobile hotspots. They evaluated different location recommenders against their defined fairness criteria and found that fairness depends on the specific fairness concerns being considered.

To mitigate the challenges of multi-scenario modeling and data fairness in the field of travel marketing, Shen et al. (2021) developed a model called the Scenario-Aware Ranking Network (SAR-Net). This model utilizes two specific attention modules to learn different scenarios by studying users' cross-scenario interactions. The proposed model was tested on Alibaba's travel marketing platform, resulting in a 5% increase in its clickthrough rate. They further suggest that this model can be applied to various travel scenarios to generate personalized and unbiased recommendations.

Wu et al. (2021) developed a two-sided Fairness-Aware Recommendation Model (TFROM) that utilizes post-processing heuristic algorithms to optimize for both C-Fairness and I-Fairness. The effectiveness of TFROM was evaluated using real-world flight data from Ctrip[10], Google local dataset[11], and Amazon review dataset[12]. The results of the experiments showed that TFROM provides better two-sided fairness while still having a minimal trade-off in personalization compared to the baseline algorithms.

Although multi-stakeholder utility models have been developed to address fairness criteria such as C-Fairness, I-Fairness, and P-Fairness, limited research has been conducted on S-Fairness, as shown in Table 2. A recent study by Merinov et al. (2022) has proposed a travel itinerary optimization approach to address S-Fairness by preventing overcrowding of tourist destinations. Their experiments were conducted on synthetic data and simulated scenarios, but further validations on real-life scenarios are required. This highlights the need for further research in this area to ensure fair recommendations for all stakeholders in actual tourism scenarios.

4.2. Multi-criteria recommendations

Several studies have been conducted on multi-objective optimization for recommendations on hotel booking platforms such as Expedia[13] (Nguyen et al., 2017) and TripAdvisor (Jannach et al., 2014; Zheng, 2017a, 2019). Even though these studies are not explicitly concerned with fairness, they can be repurposed to generate fair recommendations.

For instance, Nguyen et al. (2017) propose a learning-to-re-rank approach for solving multi-objective recommendation problems involving multiple stakeholders. They demonstrate their solution in a detailed example using an in-house Expedia dataset, integrating multistakeholder issues in hotel recommendations by incorporating consumers, platform, and provider concerns. Similarly, Zheng (2019) use multi-criteria ratings from TripAdvisor and utilize the similarity or distance between expectation and rating vectors as the utility functions to map them to different aspects such as location, room size, and cleanliness, in the case of hotel booking. They use a scoring function to recommend top-N items to the user. Jannach et al. (2014) leverage customer feedback and satisfaction analysis from TripAdvisor data and improve recommendations. Another work "CriteriaChains" by Zheng (2017a) predicts utility values one by one in a chain-like structure. Their experimental evaluation based on TripAdvisor and YahooMovies[14] rating datasets demonstrate that their proposed approach can improve the performance of multi-criteria item recommendations. The results of these studies show that these models provide improved two-sided fairness while maintaining a minimal trade-off in personalization.

10 https://www.ctrip.com
11 http://jmcauley.ucsd.edu/data/googlelocal/
12 https://jmcauley.ucsd.edu/data/amazon/
13 https://www.expedia.com/
14 https://www.yahoo.com/entertainment/movies/

TABLE 2 Summary of related works, and their main fairness criteria with an emphasis on relevant travel domain datasets (in bold).

Category	References	Dataset	Main fairness criteria			
			C-fairness	I-fairness	P-fairness	S-fairness
Fairness in TRS	Weydemann et al. (2019)	**Travel data solution**	•	•		
	Shen et al. (2021)	**Alibaba** travel marketing platform	•	•	•	
	Wu et al. (2021)	**Ctrip**, Google Local, and Amazon Review	•	•		
	Merinov et al. (2022)	synthetic data	•	○		○
	Rahmani et al. (2022b)	**Yelp**, **Gowalla**	•	•	•	
Multi-Criteria recommendations	Jannach et al. (2014)	**TripAdvisor**, **HRS.com**, YahooMovies	•	○		
	Nguyen et al. (2017)	**Expedia**	•	•	•	
	(Zheng, 2017a, 2019)	**TripAdvisor**, YahooMovies	•		○	
Multistakeholder fairness in other domains	Burke et al. (2022)	Kiva Microloans	•	•		
	Wu et al. (2022)	MovieLens	•	•		

The • signifies the acknowledged addressing of stakeholders, while the ○ denotes a partial or incidental reference for each category.

4.3. Multistakeholder fairness in other domains

Outside the tourism domain, the topic of fairness in multistakeholder applications has received a lot of attention. While these systems differ from TRS, many of these fair RS approaches can be adapted to the tourism domain by redefining their fairness concerns.

Fairness in RS, including from a multistakeholder perspective, was surveyed by Deldjoo et al. (2022); Wang et al. (2022). The authors outline fairness definitions in recommendations and classify fairness issues from various perspectives. They also summarize the datasets and measurements used in fairness studies and present a comprehensive taxonomy of fairness methods in recommendations. We refer to their papers for an in-depth review. In this paper, we discuss additional recent studies which have not been addressed in the aforementioned work.

In the work by Wu et al. (2022), the authors propose a multi-objective optimization framework called Multi-FR for addressing the issue of multistakeholder fairness-aware recommendation. Multi-FR jointly optimizes for accuracy and fairness for both consumers and producers in an end-to-end way, resulting in a guaranteed Pareto optimal solution. The authors evaluated their model using the MovieLens dataset, but the approach can be adapted for other domains such as tourism. Another related study is the work of Burke et al. (2022), in which they introduce an innovative architecture for implementing multistakeholder fairness in recommendation systems, where fairness concerns are represented as agents in a dynamic social choice environment. They evaluated their approach on Kiva Microloans,[15] an online loan lending platform, and show that it outperforms baseline methods. Similar to the domain of tourism, where the needs of different stakeholders need to be balanced, this approach can be adopted while redefining fairness concerns.

Although multi-objective optimization appears to be a promising approach for ensuring fairness for all stakeholders, it often involves a trade-off with other criteria, such as reduced user satisfaction. This outcome is counterproductive as the primary objective of a recommender system is to recommend items that fulfill user needs. Additionally, the metrics used to measure fairness are highly dependent on the domain and context and require adaptation. Moreover, most studies evaluate their models through offline analysis using either existing or synthetic datasets. Unfortunately, they lack emphasis on user acceptance of the re-ranked or fairly recommended results. Furthermore, the use of synthetic data may not accurately reflect real-life scenarios, particularly in a dynamic domain such as travel and tourism. Consequently, future research must address these issues.

5. Challenges in fair recommender systems in tourism

Tourism is a highly dynamic and rapidly growing industry, and recommender systems have become an essential tool in helping tourists make informed decisions. However, the implementation of fair and equitable recommender systems in the tourism industry

15 https://www.kiva.org/

presents numerous challenges. The complexity of balancing the needs and preferences of multiple stakeholders, such as tourists, service providers, and platform providers, creates a complex decision-making environment. Additionally, factors such as changing contexts and the diversity of domains add to the complexity of the problem.

In this section, we will examine the challenges associated with designing fair and balanced tourism recommender systems and explore possible solutions for mitigating these challenges. Through our examination of existing literature, we have identified these challenges, which can serve as valuable areas of focus for future research into developing fair tourism recommender systems. The section is organized as follows: we begin by examining the challenges faced by individual stakeholders in Section 5.1, then consider the trade-offs between different stakeholder groups in Section 5.2, explore how explanations can enhance user interfaces and transparency in Section 5.3, and conclude by addressing the shortage of publicly available data, metrics, and evaluation approaches in Section 5.

5.1. Modeling individual stakeholder utilities

In the tourism industry, modeling utilities for each stakeholder is essential, similar to other recommendation domains. However, utility modeling in tourism is a complicated process, as it is often influenced by dynamic factors such as context, seasonality, travel regulations, etc. The work by Zheng (2017b) effectively illustrates the difficulties encountered in a multistakeholder travel recommendation scenario. The authors emphasize the importance of considering the correlation among utilities due to dynamic factors that can impact stakeholder preferences. They note that these preferences may vary and be influenced by changing circumstances such as contextual factors or emotional states. For instance, when making a multi-criteria hotel booking recommendation on TripAdvisor.com, room size may be a crucial factor for a user when planning a family trip. A low rating on room size can directly influence the user's rating on other criteria such as "value" and overall rating of the hotel. To include these correlations in the model, researchers like Sahoo et al. (2012) have developed probabilistic recommendation algorithms based on pre-defined graphical relationships. Another proposed approach, "CriteriaChains" (Zheng, 2017a), predicts utility values one by one in a chain-like structure.

Our research has revealed that, although there has been some exploration of modeling the utilities of individual stakeholders such as consumers, providers, and platforms, there has been limited attention paid to the role that society plays in the recommendation process. In particular, the concept of sustainable tourism has been largely overlooked in recommendations, despite its importance in balancing the challenges of over and undertourism and reducing the environmental impact of tourism activities. A potential solution to combat over and undertourism could be optimizing the crowdedness of a location. Although there is some information available on Google, it is not readily accessible. Improving crowdedness modeling and increasing information on this topic will not only help mitigate overtourism but also support sustainable tourism. This highlights the need for a more comprehensive approach to the recommendation process in the tourism industry, one that takes into account the interests of all stakeholders, including society.

5.2. Complexity in inter-stakeholder relationships

Recommending items that meet the needs of multiple stakeholders is a challenging task, as it requires balancing different preferences and goals. Although research in this area has been conducted, there are relatively few studies specifically focused on TRS. Some approaches address the problem as an optimization problem (Weydemann et al., 2019; Shen et al., 2021; Wu et al., 2021), while others focus on providing transparent explanations for recommendations (Wang et al., 2022). However, the complexity of the problem is further compounded by factors such as shifting contexts and the diversity of domains (explained in subsection 5.1). As a result, solutions for multistakeholder TRS often involve making trade-offs among various optimization parameters (Rahmani et al., 2022b).

The relationships between stakeholders can be intricate and can impact their interactions and outcomes. In the tourism industry, for example, a consumer's relationship with the item provider can influence their ratings positively or negatively, leading to bias and unfairness in TRS Zheng (2017b). In Balakrishnan and Wörndl (2021), the authors highlight how one stakeholder's gain can negatively impact others in the same group, particularly in tourism recommender systems. For example, a user receiving a discount on a resort could cause another tourist to miss out, and a provider being selected by a customer could result in a loss of utility for other providers. To address this, the authors suggest using value-aware recommender systems that take into account both user and stakeholder utility gains (Pei et al., 2019; Abdollahpouri et al., 2020). As such, designing these systems in the multistakeholder context is a good starting point for a fair TRS that balances stakeholder utilities.

Temporal factors can also affect the relevance of recommendations. While context-aware recommendation models (Zheng et al., 2014, 2016) may improve the quality of recommendations, it is important to evaluate the effectiveness of multistakeholder recommendations in different contextual situations. Moreover, travel restrictions and tourism trends play an integral role in the design and implementation of TRS, as noted by Balakrishnan and Wörndl (2021). The authors emphasize that taking into account the impact of external factors can greatly enhance the benefits of a multistakeholder RS, including enhanced user satisfaction, higher conversion rates, and increased provider exposure. They categorize external influences into four groups based on duration and predictability: constant, deterministic recurrent, non-deterministic recurrent, and volatile. However, these external influences can pose significant challenges in designing fair TRS, as they are difficult to quantify and incorporate into recommendations.

5.3. Explanations to improve user interfaces

Fairness in recommendation is essential in ensuring that the recommendations generated do not favor any particular individual or group of individuals, such as consumers or providers. This can be achieved through both the fair usage of information in the recommendation process, as well as by ensuring that the recommendations themselves are fair (Zhang et al., 2020). Additionally, providing explanations or reasoning behind the recommendations can help users understand the fairness objectives of the recommender system, and potentially impact their perceptions of the fairness of the system. Explanations can also provide transparency, increase efficiency, effectiveness, and trust in the system, and ultimately lead to increased user satisfaction (Tintarev and Masthoff, 2010).

Explainability in recommender systems allows for the justification of a model's predictions and the identification of potential biases. It is an effective tool for increasing fairness in various branches of AI (Sonboli et al., 2021). Many researchers have also explored the relationship between explainability and fairness in recommender systems. For example, Abdollahi and Nasraoui (2018) argue that traditional metrics such as accuracy do not account for fairness in recommendations, and thus, explainable models are needed to achieve fairness. Similarly, Sonboli et al. (2021) suggest that it is not enough to simply claim a system is fair, rather, fairness goals should be effectively explained to users for them to perceive the recommender system as fair. Another study by Elahi et al. (2021) uses the Universal Design for Learning (UDL) framework to introduce three metrics for evaluating user-perceived fairness in recommender systems: Engagement, Representation, and Action & Expression; and suggests that explanations can contribute to fairness in the representation of recommendations.

By offering explanations for recommendations, transparency, trust, and user satisfaction can be promoted, and users can make more informed decisions. Despite the importance of this issue, research on how to explain recommendations with a multistakeholder fairness objective in the tourism industry is limited.

5.4. Insufficient data, missing metrics, and evaluation

The study of fairness in TRS is an emerging field, but a lack of publicly available data hinders its progress. Many studies in this area have used synthetic datasets (Merinov et al., 2022) or data from specific platforms that are not publicly accessible. For example, some have used an in-house dataset from Expedia, which is not available to the general public (Nguyen et al., 2017). Additionally, the available datasets often lack essential information such as user interactions or preferences and typically contain only limited fairness-related metadata such as gender and age. Moreover, to address environmental impact and incorporate societal perspectives into the recommendation process, it is essential to have access to data that quantifies metrics such as environmental impact and the crowdedness of a place. This makes it difficult to reproduce results or generalize findings. The data availability problem is also present in the music domain, as noted by Dinnissen and Bauer (2022). Although there is debate on whether such data should be made publicly available, it is clear that more representative and detailed data is needed to develop effective and fair TRS.

The fairness metrics used in fair TRS research are highly specific to a particular domain or context, making it challenging to generalize their application. The complexity of modeling utilities in the tourism domain, as discussed in Section 5.1, further complicates the issue. Despite the advances made by recent methods like Bauer et al. (2023), which provide researchers with tools for carrying out, analyzing, and comprehending recommendation computations through the use of 5 datasets, 11 metrics, and 21 recommendation algorithms, it is not possible to extend its results to all scenarios.

Furthermore, achieving fairness in recommendation systems is a complex task due to the societal construct of fairness, with various definitions existing (Narayanan, 2018). To address this, several researchers have proposed different ways of operationalizing fairness constraints. However, many of these approaches lack evidence or argumentation justifying the chosen fairness metrics' practical relevance in general or specific application settings (Jannach and Abdollahpouri, 2023). Some prior works, including Abdollahpouri et al. (2019a), have loosely associated fair recommendations with reducing popularity bias by matching with a target distribution or metric threshold. However, as pointed out by Jannach and Abdollahpouri (2023) it remains unclear what normative claim justifies recommending less popular items, which could be of poor quality and perceived as unfair by users. Moreover, recommending mostly popular items may negatively impact accuracy and affect different user groups in distinct ways, as shown in previous studies on movies (Abdollahpouri et al., 2019b) and music (Kowald et al., 2020) domains.

While most studies evaluate their models through offline analysis or using existing datasets, there is a lack of focus on user acceptance of the re-ranked or fair recommended results. This is a vital aspect of recommender systems, as they must not only align with user preferences but also be fair to all stakeholders. Future research should prioritize this aspect to ensure the practicality and effectiveness of the models developed.

6. Conclusion

In recent years, the prevalence of unfairness in recommender systems has become a topic of increasing concern, leading to the development of various definitions, metrics, and techniques to promote fairness. As a multi-faceted concept, ensuring fairness in recommender systems involves addressing the needs of multiple stakeholders, both within and beyond the system. This paper reviews existing literature on fairness in tourism recommender systems, categorizes stakeholders based on their primary fairness criteria and discusses the challenges associated with developing fair recommender systems.

While research has been done on fairness in RS in other domains, the domain of travel and tourism remains largely unexplored. The majority of studies in the tourism sector have centered on fairness in accommodation and restaurant recommendations, while other areas such as fair trip planning and transportation have received limited attention. Additionally, there has been limited research into integrating societal concerns as a stakeholder when defining the utility function of a recommendation. Future work should prioritize balancing the requirements of society with those of the other stakeholders.

Author contributions

AB: literature analysis. PB: literature analysis, contributed to Sections 3 and 4. WW: supervised the project. All authors contributed to the manuscript revision, read, and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdollahi, B., and Nasraoui, O. (2018). "Transparency in fair machine learning: the case of explainable recommender systems," in *Human and Machine Learning. Human—Computer Interaction Series*, eds J. Zhou, and F. Chen (Cham: Springer). p. 21–35. doi: 10.1007/978-3-319-90403-0_2

Abdollahpouri, H. (2019). "Popularity bias in ranking and recommendation," in *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. p. 529–530. doi: 10.1145/3306618.3314309

Abdollahpouri, H., Adomavicius, G., Burke, R., Guy, I., Jannach, D., Kamishima, T., et al. (2020). Multistakeholder recommendation: survey and research directions. *User Model. User-Adapt. Interact.* 30, 127–158. doi: 10.1007/s11257-019-09256-1

Abdollahpouri, H., and Burke, R. (2019). "Multi-stakeholder recommendation and its connection to multi-sided fairness," in *Proc. RMSE Workshop held in Conjunction With the 13th ACM Conference on Recommender Systems (RecSys'19)* (Copenhagen).

Abdollahpouri, H., and Burke, R. (2021). "Multistakeholder recommender systems," in *Recommender systems handbook*. Berlin, Germany: Springer. p. 647–677. doi: 10.1007/978-1-0716-2197-4_17

Abdollahpouri, H., Burke, R., and Mobasher, B. (2017). "Controlling popularity bias in learning-to-rank recommendation," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*. p. 42–46. doi: 10.1145/3109859.3109912

Abdollahpouri, H., Burke, R., and Mobasher, B. (2019a). Managing popularity bias in recommender systems with personalized re-ranking. *arXiv*. doi: 10.48550/arXiv.1901.07555

Abdollahpouri, H., and Mansoury, M. (2020). "Multi-sided exposure bias in recommendation. In International Workshop on Industrial Recommendation Systems (IRS2020)," in *Conjunction with ACM KDD 2020* (Virtual).

Abdollahpouri, H., Mansoury, M., Burke, R., and Mobasher, B. (2019b). "The unfairness of popularity bias in recommendation," in *Proc. RMSE workshop held in conjunction with the 13th ACM Conference on Recommender Systems (RecSys'19)* (Copenhagen).

Baeza-Yates, R. (2018). Bias on the web. *Commun. ACM*. 61, 54–61. doi: 10.1145/3209581

Balakrishnan, G., and Wörndl, W. (2021). "Multistakeholder recommender systems in tourism," in *Proc. Workshop on Recommenders in Tourism (RecTour 2021)* (Amsterdam).

Banerjee, A., Patro, G. K., Dietz, L. W., and Chakraborty, A. (2020). "Analyzing 'Near Me' services: Potential for exposure bias in location-based retrieval," in *2020 IEEE International Conference on Big Data (Big Data)* (Atlanta). p. 3642–3651. doi: 10.1109/BigData50022.2020.9378476

Bauer, C., Chung, L., Cornelissen, A., van Driessel, I., van der Hoorn, D., Jong, Y., et al. (2023). "Fairreckit: A web-based analysis software for recommender evaluations," in *8th ACM SIGIR Conference on Human Information Interaction and Retrieval, CHIIR 2023*. New York, NY, USA: ACM. doi: 10.1145/3576840.3578274

Bellogín, A., Castells, P., and Cantador, I. (2017). Statistical biases in information retrieval metrics for recommender systems. *Inf. Retrieval J.* 20, 606–634. doi: 10.1007/s10791-017-9312-z

Beutel, A., Chen, J., Doshi, T., Qian, H., Wei, L., Wu, Y., et al. (2019). "Fairness in recommendation ranking through pairwise comparisons," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK). p. 2212–2220. doi: 10.1145/3292500.3330745

Biega, A. J., Gummadi, K. P., and Weikum, G. (2018). "Equity of attention: Amortizing individual fairness in rankings," in *The 41st international acm sigir conference on research & development in information retrieval* (Ann Arbor Michigan). p. 405–414. doi: 10.1145/3209978.3210063

Boratto, L., Fenu, G., and Marras, M. (2021). Interplay between upsampling and regularization for provider fairness in recommender systems. *User Model. User-Adapted Interact.* 31, 421–455. doi: 10.1007/s11257-021-09294-8

Bouveret, S., Chevaleyre, Y., Maudet, N., and Moulin, H. (2016). *Fair Allocation of Indivisible Goods*. p. 284–310. doi: 10.1017/CBO9781107446984.013

Buet-Golfouse, F., and Utyagulov, I. (2022). "Towards fair multi-stakeholder recommender systems," in *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (Barcelona). p. 255–265. doi: 10.1145/3511047.3538031

Burke, R. (2017). "Multisided fairness for recommendation," in *Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2017)*.

Burke, R., Mattei, N., Grozin, V., Voida, A., and Sonboli, N. (2022). "Multi-agent social choice for dynamic fairness-aware recommendation," in *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (Barcelona). p. 234–244. doi: 10.1145/3511047.3538032

Burke, R., Sonboli, N., and Ordonez-Gauger, A. (2018). "Balanced neighborhoods for multi-sided fairness in recommendation," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, volume 81 of Proceedings of Machine Learning Research* (New York, NY: PMLR). p. 202-214.

Burke, R. D., Abdollahpouri, H., Mobasher, B., and Gupta, T. (2016). "Towards multi-stakeholder utility evaluation of recommender systems," in *UMAP (Extended Proceedings)* 750.

Camarda, D., and Grassini, L. (2003). *Environmental impacts of tourism. Conference on the Relationships between Global Trades and Local Resources in the Mediterranean Region*. Morocco: Rabat. p. 263–270.

Castillo, M. Z. F. B. C., Megahed, S. H. M., and Baeza-Yates, R. (2017). "Fa* ir: A fair top-k ranking algorithm," in *ACM CIKM. Meike Zehlike Francesco Bonchi Carlos Castillo Sara Hajian Mohamed Megahed and Ricardo Baeza-Yates* (Turin).

Celis, L. E., Straszak, D., and Vishnoi, N. K. (2018). "Ranking with fairness constraints," in *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018), volume 107*, Chatzigiannakis, I., Kaklamanis, C., Marx, D., and Sannella, D., (eds). Dagstuhl, Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Chuklin, A., Markov, I., and Rijke, M. (2015). Click models for web search. *Synth. Lect. Inf. Concepts Retr.* 7, 1–115. doi: 10.1007/978-3-031-02294-4

Ciampaglia, G. L., Nematzadeh, A., Menczer, F., and Flammini, A. (2018). How algorithmic popularity bias hinders or promotes quality. *Sci. Rep.* 8, 15951. doi: 10.1038/s41598-018-34203-2

Craswell, N., Zoeter, O., Taylor, M., and Ramsey, B. (2008). "An experimental comparison of click position-bias models," in *Proceedings of the 2008 international conference on web search and data mining* (Palo Alto California). p. 87–94. doi: 10.1145/1341531.1341545

Dastgerdi, A. S., and De Luca, G. (2023). A post-pandemic look at short-term tourist accommodations and world heritage site management. *J. Tour. Cultural Change*. 1–12. doi: 10.1080/14766825.2022.2164502

Deldjoo, Y., Jannach, D., Bellogin, A., Difonzo, A., and Zanzonelli, D. (2022). A survey of research on fair recommender systems. *arXiv*. doi: 10.48550/arXiv.2205.11127

Delic, A., Neidhardt, J., Nguyen, T. N., and Ricci, F. (2018). An observational user study for group recommender systems in the tourism domain. *Inf. Techn. Tour*. 19, 87–116. doi: 10.1007/s40558-018-0106-y

Dinnissen, K., and Bauer, C. (2022). Fairness in music recommender systems: a stakeholder-centered mini review. *Front. big Data*. 5, 913608. doi: 10.3389/fdata.2022.913608

Dodds, R., and Butler, R. (2019). The phenomena of overtourism: a review. *Int. J Tour. Cities*. 5, 519–528. doi: 10.1515/9783110607369

Draws, T., Tintarev, N., Gadiraju, U., Bozzon, A., and Timmermans, B. (2021). "This is not what we ordered: Exploring why biased search result rankings affect user attitudes on debated topics," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. p. 295–305. doi: 10.1145/3404835.3462851

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. (2012). "Fairness through awareness," in *Proceedings of the 3rd innovations in theoretical computer science conference* (Cambridge, MA), p. 214–226. doi: 10.1145/2090236.2090255

Edelman, B., Luca, M., and Svirsky, D. (2017). Racial discrimination in the sharing economy: Evidence from a field experiment. *Am. Econ. J*. 9, 1–22. doi: 10.1257/app.20160213

Ekstrand, M., Tian, M., Azpiazu, I., Ekstrand, J., Anuyah, O., McNeill, D., et al. (2018). "All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness," in *Conference on Fairness, Accountability and Transparency* (New York, NY). p. 172–186.

Ekstrand, M. D., Das, A., Burke, R., and Diaz, F. (2020). "Fairness in recommender systems," in *Recommender Systems Handbook*. p. 679. doi: 10.1007/978-1-0716-2197-4_18

Elahi, M., Abdollahpouri, H., Mansoury, M., and Torkamaan, H. (2021). "Beyond algorithmic fairness in recommender systems," in *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '21*. New York, NY, USA: Association for Computing Machinery. p. 41–46. doi: 10.1145/3450614.3461685

Fortunato, S., Flammini, A., Menczer, F., and Vespignani, A. (2006). Topical interests and the mitigation of search engine bias. *Proc. Natl. Acad. Sci. U.S.A*. 103, 12684–12689. doi: 10.1073/pnas.0605525103

Fu, Z., Xian, Y., Geng, S., De Melo, G., and Zhang, Y. (2021). "Popcorn: Human-in-the-loop popularity debiasing in conversational recommender systems," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (Virtual Event Queensland). p. 494–503. doi: 10.1145/3459637.3482461

GalíEspelt, N. (2022). Impacts of covid-19 on local tour guides. *J. Tour. Cultural Chang*. doi: 10.1080/14766825.2022.2046016

Geyik, S. C., Ambler, S., and Kenthapadi, K. (2019). "Fairness-aware ranking in search & recommendation systems with application to linkedin talent search," in *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining* (Anchorage, AK). p. 2221–2231. doi: 10.1145/3292500.3330691

Gössling, S. (2017). Tourism, information technologies and sustainability: an exploratory review. *J. Sust. Tour*. 25, 1024–1041. doi: 10.1080/09669582.2015.1122017

Gowreesunkar, V. G., and Vo Thanh, T. (2020). "Between overtourism and under-tourism: Impacts, implications, and probable solutions," in *Overtourism: Causes, implications and solutions*. p. 45–68. doi: 10.1007/978-3-030-42458-9_4

Grbovic, M., and Cheng, H. (2018). "Real-time personalization using embeddings for search ranking at airbnb," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. p. 311–320. doi: 10.1145/3219819.3219885

Gunawardena, D., and Sarathchandra, K. (2020). "Bestdish: a digital menu and food item recommendation system for restaurants in the hotel sector," in *2020 International Conference on Image Processing and Robotics (ICIP)*. p. 1-7. IEEE. doi: 10.1109/ICIP48927.2020.9367357

Gupta, A., Johnson, E., Payan, J., Roy, A. K., Kobren, A., Panda, S., et al. (2021a). "Online post-processing in rankings for fair utility maximization," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining, WSDM '21*. New York, NY, USA: Association for Computing Machinery. p. 454–462. doi: 10.1145/3437963.3441724

Gupta, A., Yadav, R., Nair, A., Chakraborty, A., Ranu, S., and Bagchi, A. (2022). Fairfoody: Bringing in fairness in food delivery. *Proc. Innov. Appl. Artif. Intell. Conf*. 36, 11900–11907. doi: 10.1609/aaai.v36i11.21447

Gupta, S., Wang, H., Lipton, Z., and Wang, Y. (2021b). "Correcting exposure bias for link recommendation," in *International Conference on Machine Learning* (Virtual). p. 3953–3963.

Hao, F., Xiao, Q., and Chon, K. (2020). Covid-19 and china" hotel industry: Impacts, a disaster management framework, and post-pandemic agenda. *Int. J. Hospital. Manag*. 90, 102636. doi: 10.1016/j.ijhm.2020.102636

Hardt, M., Price, E., and Srebro, N. (2016). "Equality of opportunity in supervised learning," in *Advances in neural information processing systems* (Barcelona). p. 29.

Herzog, D., and Wörndl, W. (2019). "A user study on groups interacting with tourist trip recommender systems in public spaces," in *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '19*. New York, NY, USA: Association for Computing Machinery. p. 130–138. doi: 10.1145/3320435.3320449

Higgins-Desbiolles, F., Carnicelli, S., Krolikowski, C., Wijesinghe, G., and Boluk, K. (2019). Degrowing tourism: rethinking tourism. *J. Sust. Tour*. 27, 1926–1944. doi: 10.1080/09669582.2019.1601732

Hospers, G.-J. (2019). Overtourism in European cities: from challenges to coping strategies. *CESifo Forum*. 20, 20–24.

Huang, H., Gartner, G., Krisp, J. M., Raubal, M., and Van de Weghe, N. (2018). Location based services: ongoing evolution and research agenda. *J. Location Based Serv*. 12, 63–93. doi: 10.1080/17489725.2018.1508763

Isinkaye, F., Folajimi, Y., and Ojokoh, B. (2015). Recommendation systems: principles, methods and evaluation. *Egyptian Inf. J*. 16, 261–273. doi: 10.1016/j.eij.2015.06.005

Jaeger, B., and Sleegers, W. (2020). Racial discrimination in the sharing economy: Evidence from airbnb markets across the world. *Am. Econ. J*. 9, 1–20. doi: 10.31234/osf.io/qusxf

Jannach, D., and Abdollahpouri, H. (2023). A survey on multi-objective recommender systems. *Front. Big Data*. 6, 24. doi: 10.3389/fdata.2023.1157899

Jannach, D., and Bauer, C. (2020). Escaping the mcnamara fallacy: toward more impactful recommender systems research. *Ai Magazine*. 41, 79–95. doi: 10.1609/aimag.v41i4.5312

Jannach, D., Lerche, L., Kamehkhosh, I., and Jugovac, M. (2015). What recommenders recommend: an analysis of recommendation biases and possible countermeasures. *User Model. User-adapt Interact*. 25, 427–491. doi: 10.1007/s11257-015-9165-3

Jannach, D., Zanker, M., and Fuchs, M. (2014). Leveraging multi-criteria customer feedback for satisfaction analysis and improved recommendations. *Inf. Technol. Tour*. 14, 119–149. doi: 10.1007/s40558-014-0010-z

Kamishima, T., Akaho, S., Asoh, H., and Sakuma, J. (2013). "Efficiency improvement of neutrality-enhanced recommendation," in *Decisions@ RecSys*. p. 1–8. Princeton, New Jersey, US: Citeseer.

Kamishima, T., Akaho, S., Asoh, H., and Sakuma, J. (2014). "Correcting popularity bias by enhancing recommendation neutrality," in *RecSys Posters* p. 805.

Kamishima, T., Akaho, S., Asoh, H., and Sakuma, J. (2018). "Recommendation independence," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, volume 81 of Proceedings of Machine Learning Research*. p. 187-201. PMLR.

Kangas, I., Schwoerer, M., and Bernardi, L. J. (2021). "Recommender systems for personalized user experience: Lessons learned at booking.com," in *Proceedings of the 15th ACM Conference on Recommender Systems, RecSys '21*. New York, NY, USA: Association for Computing Machinery. p. 583–586. doi: 10.1145/3460231.3474611

Khenissi, S., and Nasraoui, O. (2020). Modeling and counteracting exposure bias in recommender systems. *arXiv*. doi: 10.48550/arXiv.2001.04832

Kokkodis, M., and Lappas, T. (2020). Your hometown matters: popularity-difference bias in online reputation platforms. *Inf. Syst. Res*. 31, 412–430. doi: 10.1287/isre.2019.0895

Kowald, D., Schedl, M., and Lex, E. (2020). "The unfairness of popularity bias in music recommendation: a reproducibility study," in *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Proceedings, Part II 42*. p. 35-42. Lisbon, Portugal: Springer. doi: 10.1007/978-3-030-45442-5_5

Lahoti, P., Gummadi, K. P., and Weikum, G. (2019). "ifair: Learning individually fair data representations for algorithmic decision making," in *2019 ieee 35th international conference on data engineering (icde)*. New York City: IEEE. p. 1334–1345. doi: 10.1109/ICDE.2019.00121

Li, Y. (2020). Handling position bias for unbiased learning to rank in hotels search. *arXiv*. doi: 10.48550/arXiv.2002.12528

Lin, A., Wang, J., Zhu, Z., and Caverlee, J. (2022). "Quantifying and mitigating popularity bias in conversational recommender systems," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management* (Atlanta, GA). p. 1238–1247. doi: 10.1145/3511808.3557423

Liu, W., and Burke, R. (2018). "Personalizing fairness-aware re-ranking," in *2nd FATREC Workshop on Responsible Recommendation* (Vancouver, CA).

Liu, W., Guo, J., Sonboli, N., Burke, R., and Zhang, S. (2019). "Personalized fairness-aware re-ranking for microlending," in *Proceedings of the 13th ACM Conference on Recommender Systems*. p. 467–471. doi: 10.1145/3298689.3347016

Mansoury, M., Mobasher, B., Burke, R., and Pechenizkiy, M. (2019). "Bias disparity in collaborative recommendation: Algorithmic evaluation and comparison," in *Workshop on Recommendation in Multi-Stakeholder Environments (RMSE) at ACM RecSys 2019* (Copenhagen).

Masthoff, J., and Delić, A. (2012). "Group recommender systems: Beyond preference aggregation," in *Recommender Systems Handbook*. Berlin, Germany: Springer. p. 381–420. doi: 10.1007/978-1-0716-2197-4_10

Mavridis, T., Hausl, S., Mende, A., and Pagano, R. (2020). "Beyond algorithms: Ranking at scale at booking," in *ComplexRec-ImpactRS@ RecSys* (Virtual).

Merinov, P., Massimo, D., and Ricci, F. (2022). "Sustainability driven recommender systems," in *CEUR Workshop Proceedings*. p. 3177.

Narayanan, A. (2018). "Translation tutorial: 21 fairness definitions and their politics," in *Proc. conf. fairness accountability transp., new york, usa*. p. 3.

Nguyen, P., Dines, J., and Krasnodebski, J. (2017). "A multi-objective learning to re-rank approach to optimize online marketplaces for multiple stakeholders," in *Workshop on Value-Aware and Multistakeholder Recommendation*.

Pachot, A., Albouy-Kissi, A., Albouy-Kissi, B., and Chausse, F. (2021). "Multiobjective recommendation for sustainable production systems," in *MORS workshop held in conjunction with the 15th ACM Conference on Recommender Systems (RecSys)* 2021 (Amsterdam).

Pala, U. (2021). Popularity bias of online hotel reviews: An alternative evaluation using sentiment analysis. in *Proceedings of the Global Conference on Services Management: Volume 2*. p. 101. doi: 10.5038/2641-5062-v2

Park, Y.-J., and Tuzhilin, A. (2008). "The long tail of recommender systems and how to leverage it," in *Proceedings of the 2008 ACM conference on Recommender systems* (Lausanne). p. 11–18. New York, NY: ACM. doi: 10.1145/1454008.1454012

Patro, G. K., Biswas, A., Ganguly, N., Gummadi, K. P., and Chakraborty, A. (2020a). "Fairrec: Two-sided fairness for personalized recommendations in two-sided platforms," in *Proceedings of The Web Conference 2020* (Taipei). p. 1194–1204. doi: 10.1145/3366423.3380196

Patro, G. K., Chakraborty, A., Banerjee, A., and Ganguly, N. (2020b). "Towards safety and sustainability: designing local recommendations for post-pandemic world," in *Fourteenth ACM conference on recommender systems*. p. 358–367. doi: 10.1145/3383313.3412251

Pedreshi, D., Ruggieri, S., and Turini, F. (2008). "Discrimination-aware data mining," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (Las Vegas, NV). p. 560–568. doi: 10.1145/1401890.1401959

Pei, C., Yang, X., Cui, Q., Lin, X., Sun, F., Jiang, P., et al. (2019). "Value-aware recommendation based on reinforcement profit maximization," in *The World Wide Web Conference* (San Francisco, CA). p. 3123–3129. doi: 10.1145/3308558.3313404

Rabanser, U., and Ricci, F. (2005). "Recommender systems: Do they have a viable business model in e-tourism?," in *Information and Communication Technologies in Tourism 2005* Frew, A. J., (eds). Vienna: Springer Vienna. p. 160–171. doi: 10.1007/3-211-27283-6_15

Rahmani, H. A., Deldjoo, Y., and Di Noia, T. (2022a). The role of context fusion on accuracy, beyond-accuracy, and fairness of point-of-interest recommendation systems. *Expert Syst. Appl*. 205, 117700. doi: 10.1016/j.eswa.2022.117700

Rahmani, H. A., Deldjoo, Y., Tourani, A., and Naghiaei, M. (2022b). "The unfairness of active users and popularity bias in point-of-interest recommendation," in *Advances in Bias and Fairness in Information Retrieval*, Boratto, L., Faralli, S., Marras, M., and Stilo, G. (eds). p. 56–68. doi: 10.1007/978-3-031-09316-6_6

Ranjbar Kermany, N., Zhao, W., Yang, J., Wu, J., and Pizzato, L. (2021). A fairness-aware multi-stakeholder recommender system. *World Wide Web*. 24, 1995–2018. doi: 10.1007/s11280-021-00946-8

Rastegarpanah, B., Gummadi, K. P., and Crovella, M. (2019). "Fighting fire with fire: Using antidote data to improve polarization and fairness of recommender systems," in *Proceedings of the twelfth ACM international conference on web search and data mining* (Melbourne, VIC). p. 231–239. doi: 10.1145/3289600.3291002

Ricci, F., Rokach, L., and Shapira, B. (2020). "Introduction to recommender systems handbook," in *Recommender systems handbook*. p. 1–35. Berlin: Springer. doi: 10.1007/978-1-0716-2197-4_1

Romei, A., and Ruggieri, S. (2014). A multidisciplinary survey on discrimination analysis. *Knowl. Engineering Rev*. 29, 582–638. doi: 10.1017/S0269888913000039

Sahoo, N., Krishnan, R., Duncan, G., and Callan, J. (2012). Research note "The halo effect in multicomponent ratings and its implications for recommender systems: the case of yahoo! movies. *Infor. Syst. Res*. 23, 231–246. doi: 10.1287/isre.1100.0336

Serbos, D., Qi, S., Mamoulis, N., Pitoura, E., and Tsaparas, P. (2017). "Fairness in package-to-group recommendations," in *Proceedings of the 26th International Conference on World Wide Web, WWW '17* (Perth). Republic and Canton of Geneva: CHE. International World Wide Web Conferences Steering Committee. p. 371–379. doi: 10.1145/3038912.3052612

Shen, Q., Tao, W., Zhang, J., Wen, H., Chen, Z., and Lu, Q. (2021). "Sar-net: A scenario-aware ranking network for personalized fair recommendation in hundreds of travel scenarios," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (Virtual Event, QLD). p. 4094–4103. doi: 10.1145/3459637.3481948

Singh, A., and Joachims, T. (2018). "Fairness of exposure in rankings," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (London). p. 2219–2228. doi: 10.1145/3219819.3220088

Sonboli, N., Smith, J. J., Cabral Berenfus, F., Burke, R., and Fiesler, C. (2021). "Fairness and transparency in recommendation: The users perspective," in *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (Utrecht). p. 274–279. doi: 10.1145/3450613.3456835

Speicher, T., Heidari, H., Grgic-Hlaca, N., Gummadi, K. P., Singla, A., Weller, A., et al. (2018). "A unified approach to quantifying algorithmic unfairness: Measuring individual &group unfairness via inequality indices," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (London). p. 2239–2248. doi: 10.1145/3219819.3220046

Steck, H. (2018). "Calibrated recommendations," in *Proceedings of the 12th ACM conference on recommender systems* (Vancouver British Columbia, CA). p. 154–162. doi: 10.1145/3240323.3240372

Sühr, T., Biega, A. J., Zehlike, M., Gummadi, K. P., and Chakraborty, A. (2019). "Two-sided fairness for repeated matchings in two-sided markets: A case study of a ride-hailing platform," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK). p. 3082–3092. doi: 10.1145/3292500.3330793

Tacli, Y., Yalcin, E., and Bilge, A. (2022). "Novel approaches to measuring the popularity inclination of users for the popularity bias problem," in *2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. New York: IEEE. p. 555–560. doi: 10.1109/ISMSIT56059.2022.9932663

Tintarev, N., and Masthoff, J. (2010). "Designing and evaluating explanations for recommender systems," in *Recommender systems handbook*. Berlin: Springer. p. 479–510. doi: 10.1007/978-0-387-85820-3_15

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Sci*. 37, 530–552. doi: 10.1287/mksc.2017.1072

Wang, Y., Ma, W., Zhang, M., Liu, Y., and Ma, S. (2022). A survey on the fairness of recommender systems. *ACM J*. 47, 1–43. doi: 10.1145/3547333

Wei, T., Feng, F., Chen, J., Wu, Z., Yi, J., and He, X. (2021). "Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. p. 1791–1800. doi: 10.1145/3447548.3467289

Weydemann, L., Sacharidis, D., and Werthner, H. (2019). "Defining and measuring fairness in location recommendations," in *Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based recommendations, geosocial networks and geoadvertising* (Chicago, IL). p. 1-8. doi: 10.1145/3356994.3365497

Wu, H., Ma, C., Mitra, B., Diaz, F., and Liu, X. (2022). A multi-objective optimization framework for multi-stakeholder fairness-aware recommendation. *ACM Trans Inf Syst*. 41, 1–29. doi: 10.1145/3564285

Wu, Y., Cao, J., Xu, G., and Tan, Y. (2021). "Tfrom: A two-sided fairness-aware recommendation model for both customers and providers," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 1013–1022. doi: 10.1145/3404835.3462882

Xiao, L., Min, Z., Yongfeng, Z., Zhaoquan, G., Yiqun, L., and Shaoping, M. (2017). "Fairness-aware group recommendation with pareto-efficiency," in *Proceedings of the Eleventh ACM Conference on Recommender System* (Como). p. 107–115. doi: 10.1145/3109859.3109887

Yalcin, E., and Bilge, A. (2022). Evaluating unfairness of popularity bias in recommender systems: a comprehensive user-centric analysis. *Infor. Proces. Manag*. 59, 103100. doi: 10.1016/j.ipm.2022.103100

Yang, K., and Stoyanovich, J. (2017). "Measuring fairness in ranked outputs," in *Proceedings of the 29th international conference on scientific and statistical database management*. p. 1–6. doi: 10.1145/3085504.3085526

Yang, Y., Li, M., Hu, X., Pan, G., Huang, W., Wang, J., et al. (2021). "Exploring exposure bias in recommender systems from causality perspective," in *2021 IEEE 21st International Conference on Software Quality, Reliability and Security Companion (QRS-C)* (Hainan Island). p. 425–432. doi: 10.1109/QRS-C55045.2021.00069

Yao, S., and Huang, B. (2017). Beyond parity: fairness objectives for collaborative filtering. *Adv. Inf. Processing Syst*. p. 30.

Zafar, M. B., Valera, I., Gomez Rodriguez, M., and Gummadi, K. P. (2017). "Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment," in *Proceedings of the 26th international conference on world wide web* (Perth). p. 1171–1180. doi: 10.1145/3038912.3052660

Zemel, R., Wu, Y., Swersky, K., Pitassi, T., and Dwork, C. (2013). "Learning fair representations," in *International conference on machine learning*. New York City, NY, USA: PMLR. p. 325–333.

Zhang, L., Xiong, S., Zhang, L., Bai, L., and Yan, Q. (2022). Reducing racial discrimination in the sharing economy: empirical results from airbnb. *Int. J. Hospital. Manag*. 102, 103151. doi: 10.1016/j.ijhm.2022.103151

Zhang, Y., Chen, X., et al. (2020). Explainable recommendation: A survey and new perspectives. *Trends Inf. Retr*. 14, 1–101. doi: 10.1561/9781680836592

Zheng, Y. (2017a). "Criteria chains: a novel multi-criteria recommendation approach," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. p. 29–33. doi: 10.1145/3025171.3025215

Zheng, Y. (2017b). "Multi-stakeholder recommendation: Applications and challenges," in *Workshop on Value-Aware and Multistakeholder Recommendation*.

Zheng, Y. (2019). "Utility-based multi-criteria recommender systems," in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*. p. 2529–2531. doi: 10.1145/3297280.3297641

Zheng, Y., Mobasher, B., and Burke, R. (2014). "Cslim: Contextual slim recommendation algorithms," in *Proceedings of the 8th ACM Conference on Recommender Systems*. p. 301–304. doi: 10.1145/2645710.2645756

Zheng, Y., Mobasher, B., and Burke, R. (2016). "User-oriented context suggestion," in *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*. p. 249–258. doi: 10.1145/2930238.2930252

Zhou, J. P., Cheng, Z., Pérez, F., and Volkovs, M. (2020). "Tafa: Two-headed attention fused autoencoder for context-aware recommendations," in *Fourteenth ACM Conference on Recommender Systems*. p. 338–347. doi: 10.1145/3383313.3412268

Zhu, Z., Wang, J., and Caverlee, J. (2020). "Measuring and mitigating item under-recommendation bias in personalized ranking systems," in *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. p. 449–458. doi: 10.1145/3397271.3401177

frontiers | Frontiers in Big Data

TYPE Mini Review
PUBLISHED 10 August 2023
DOI 10.3389/fdata.2023.1239705

Check for updates

OPEN ACCESS

EDITED BY
Emanuel Lacić,
Know Center, Austria

REVIEWED BY
Tomislav Duricic,
Graz University of Technology, Austria
Sead Delalic,
University of Sarajevo, Bosnia and Herzegovina

*CORRESPONDENCE
Benedikt Loepp
✉ benedikt.loepp@uni-due.de

RECEIVED 13 June 2023
ACCEPTED 31 July 2023
PUBLISHED 10 August 2023

CITATION
Loepp B (2023) Multi-list interfaces for recommender systems: survey and future directions. *Front. Big Data* 6:1239705.
doi: 10.3389/fdata.2023.1239705

COPYRIGHT
© 2023 Loepp. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Multi-list interfaces for recommender systems: survey and future directions

Benedikt Loepp*

Department of Computer Science and Applied Cognitive Science, University of Duisburg-Essen, Duisburg, Germany

For a long time, recommender systems presented their results in the form of simple item lists. In recent years, however, multi-list interfaces have become the de-facto standard in industry, presenting users with numerous collections of recommendations, one below the other, each containing items with common characteristics. Netflix's interface, for instance, shows movies from certain genres, new releases, and lists of curated content. Spotify recommends new songs and albums, podcasts on specific topics, and what similar users are listening to. Despite their popularity, research on these so-called "carousels" is still limited. Few authors have investigated how to simulate the user behavior and how to optimize the recommendation process accordingly. The number of studies involving users is even smaller, with sometimes conflicting results. Consequently, little is known about how to design carousel-based interfaces for achieving the best user experience. This mini review aims to organize the existing knowledge and outlines directions that may improve the multi-list presentation of recommendations in the future.

KEYWORDS

recommender systems, multi-list recommendation, carousels, user interfaces, user experience, choice overload, survey

1. Introduction

Recommender systems (RS) play a vital role in a variety of domains, successfully providing users with personalized recommendations for consumer goods and entertainment media, but also for travel destinations, educational resources, people, services, and even lifestyle choices. However, the way recommendations are presented has changed significantly in recent years, especially on e-commerce and streaming platforms: While one-dimensional lists dominated for a long time, it has now become the de-facto standard to display multiple collections of recommendations. The user interfaces display these collections one below the other in a vertically scrollable list. Each row contains a number of items with a certain commonality and can be scrolled horizontally, which is why it is called a "carousel" (Bendada et al., 2020) or "shelf" (McInerney et al., 2018). Consequently, users can select items according to different contexts, rather than just from a single list optimized for a selected criterion, e.g., long-term preferences. As visible in Figure 1, Netflix shows several rows of personalized recommendations, featuring genres, popular themes, and curated content (Gomez-Uribe and Hunt, 2015). Similarly, Spotify recommends new releases, podcasts on specific topics, and songs similar users are listening to Nazari et al. (2022).

Depending on the content, different recommendation algorithms are used in the background of the carousels. Often, the systems present a corresponding label, usually as a header above the respective row. This provides a brief explanation of what is represented by the carousel, helping users identify items that match not only their general preferences, but also their current interests and situational needs. Accordingly, the carousel type can be defined based on the scheme of explanation styles proposed by Kouki et al. (2019):

- Carousels where the explanation style is *user-based* (e.g., "popular with similar viewers") contain the results of a collaborative filtering algorithm, i.e., items well received by similar users.
- The *item-based* explanation style describes carousels that contain items similar to those that the current user has rated positively in the past (e.g., "because you watched ...").
- The *content-based* explanation style uses metadata to highlight that the items are from a certain genre, star the same cast, share similar attributes, etc. (e.g., "German pop classics").
- The *social* explanation style (e.g., "played by friends") refers to preferences of peers, friends, etc.
- Global *item popularity* (e.g., "top movies in Germany") is often used for non-personalized carousels.

While all these variants are widely used in real-world applications, there is still a very limited body of literature on the presentation of recommendations in carousels. Open questions include: *which types of carousels* are preferred by which users, *how many carousels* do they want to explore, and *how many items* per carousel ensure a good decision. Moreover, while one of the main advantages of carousel-based interfaces is their ability to accommodate a variety of contexts by providing multiple sets of recommendations, it is still unclear whether this presentation format is always the most appropriate one. This is especially true when considering users with a wide range of different characteristics, aspects such as the device being used, cognitive load, and prior knowledge, as well as domain-specific requirements.

For these reasons, we aimed to organize the literature on multi-list recommender interfaces (MLRI) in this mini review. To the best of our knowledge, such a survey does not exist yet. Therefore, we systematically examined the proceedings of relevant RS and HCI conferences (e.g., RecSys, CHI, IUI, UMAP), including their workshop proceedings. We used the ACM Digital Library and Google Scholar to identify additional papers through keyword-based searches (e.g., "carousel recommendations," "multi-list recommender interfaces"). We checked the relevance of the papers based on titles and abstracts, reviewed the relevant papers in detail, and used them for further snowballing. In Section 2, we provide an overview of the resulting set of papers. From this, in Section 3, we discuss possible directions to achieve a better understanding of carousel recommendations and to improve the design of MLRI in terms of user experience.

2. Multi-list recommender interfaces: an overview

In recent years, it has been gradually recognized that algorithmic accuracy and performance are not the only factors for the success of RS (Jugovac and Jannach, 2017; Loepp et al., 2019). However, the presentation of recommendations has not received the same attention as other more user-oriented aspects. Very few authors have explored alternatives to one-dimensional lists, presenting items and arranging the user interface in different ways (Lousame and Sánchez, 2009; Nanou et al., 2010; Guntuku et al., 2016; Beel and Dixon, 2021). This seems inexplicable given the potentially strong impact of the presentation format on the user

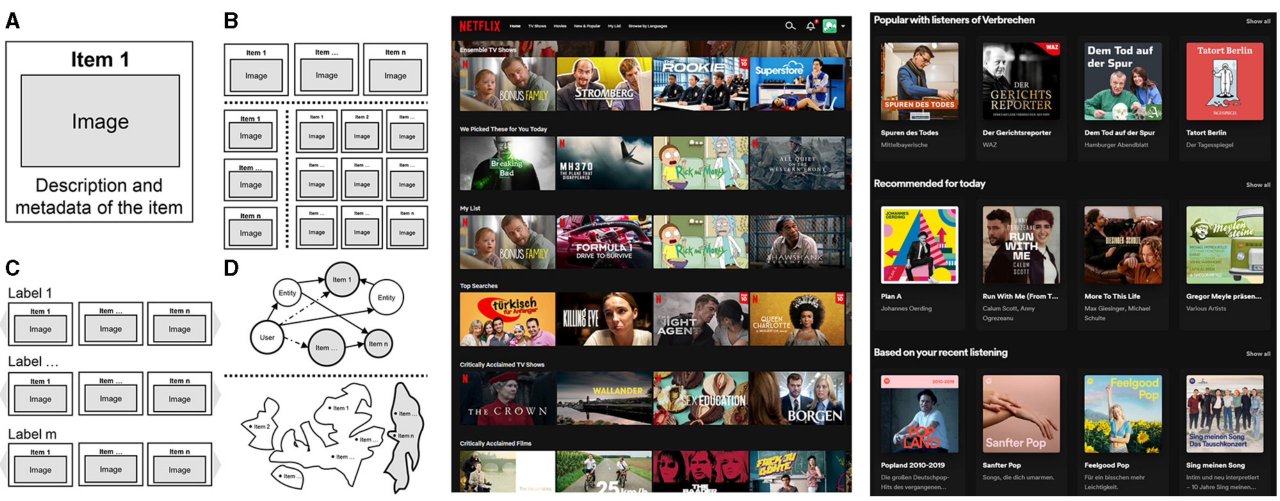

FIGURE 1
Recommender systems present their results in different ways (left), e.g., in the form of **(A)** single items or **(B)** lists of items, which may be arranged horizontally, vertically, or as a grid, but also using **(C)** carousel-based interfaces or **(D)** advanced visualization techniques such as graphs and maps. Concrete examples of carousel-based interfaces are those of Netflix (center) and Spotify (right).

experience (Knijnenburg and Willemsen, 2015). One of the most influential studies in this regard was conducted by Bollen et al. (2010). They investigated the relationship between the length of a recommendation list, the diversity of the contained items, and the occurrence of choice overload effects, and found that there is an optimal number of recommendations with respect to the balance between user satisfaction and the difficulty of making a decision. They concluded that sets of seven to ten items can be both attractive and sufficiently varied, while still being manageable for users.

The meta analysis of Scheibehenne et al. (2010) confirmed that choice overload depends on factors such as domain knowledge and decision-making strategy. However, although MLRI have become the de-facto standard in industry (cf. Section 1), most academic attempts to understand and improve the presentation of recommendations have been focused on single sets, displayed with either horizontal or vertical orientation, containing items ordered by decreasing relevance according to a selected criterion, usually long-term preferences. This also applies to the few studies in which the results were arranged as a grid. Here, the interfaces contained multiple rows, but still represented a single recommendation list, wrapped multiple times, with no option to scroll (cf. Chen and Pu, 2010; Kammerer and Gerjets, 2010).

Some of the studies on critique-based systems can be seen as exceptions, since the recommendations were displayed in groups formed on the basis of suggested critiquing options (cf. Chen and Pu, 2012a,b). However, in these cases, the purpose of item categorization was to improve the critiquing process, rather than to provide a set of diverse lists to facilitate decision making in a variety of contexts. Apart from that, the presence of categories has almost exclusively been investigated in consumer research (Knijnenburg and Willemsen, 2015). On the other hand, there exists a wide range of more advanced approaches that visualize recommendations in a more informative and appealing way than conventional lists. Numerous studies have confirmed the positive effects of graphs and maps in user-oriented dimensions such as control and transparency (He et al., 2016; Kunkel and Ziegler, 2023). However, these approaches are mostly of academic nature and too complex to be widely applied. Figure 1 provides an overview of these methods.[1]

In summary, there is a lack of research on carousel recommendations. As the next sections will show, this is especially true for questions such as those raised in Section 1, but also the impact of situational needs and individual differences. In general, personal characteristics such as expertise and decision making have not yet received much attention in RS research. The few existing studies have examined specific effects, e.g., on the preferred level of control (Jin et al., 2020), the perception of explanations (Millecamp et al., 2019), or the overall user behavior (Kleemann et al., 2021), and have always presented recommendations in a traditional way. Moreover, the literature review will show that while RS have been successful in many application scenarios (see again Section 1), research on MLRI is still limited to a few selected domains.

1 Note that there are slight differences in the definition of single- and multi-list user interfaces (cf. Jannach et al., 2021; Starke et al., 2022).

2.1. Algorithms, metrics, and models for carousel recommendations

At the same time, numerous commercial system providers have demonstrated the positive effects of carousel-based interfaces. The authors of the corresponding publications have proposed algorithmic improvements, e.g., to optimize how the collections are ordered among each other, how they are filled with items, and how labels are assigned (Wu et al., 2016, 2021; McInerney et al., 2018; Bendada et al., 2020; Lo et al., 2021). Singal et al. (2021) even investigated how to implement carousels independently of the underlying algorithms, requiring only standard user-item representations. Based on dimensionality reduction and clustering of the resulting item embeddings, this approach provides a generic way to create item collections, predict their usefulness, and find appropriate labels. Table 1 outlines the various contributions, but also shows that most of the findings stem from offline experiments and (more rarely) online A/B tests. Accordingly, metrics were used that relied purely on item clicks, largely ignoring richer behavioral data such as scrolling, responses to the mere presence of labels or items, and specific characteristics of the domain, user, or situation. Moreover, most comparisons were made against single lists, since authors (including academics) tried to optimize each collection individually (cf. Bendada et al., 2020; Jeunen and Goethals, 2021).

However, there are some notable exceptions, where the authors have attempted to model user behavior specifically for multiple carousels. Inspired by studies on search user interfaces, Felicioni et al. (2021) assumed that users follow a "golden triangle," i.e., their attention decreases from the top-left corner to the bottom and the right. From this, Ferrari Dacrema et al. (2022) formally defined an extension of the well-known NDCG metric, N2DCG, where the discounted cumulative gain g is calculated as follows:

$$2DCG_u = \sum_{j=1}^{V} \sum_{k=1}^{H} g_{ujk} d_{jk}, \tag{1}$$

With V and H representing the number of carousels (vertical) and items (horizontal), respectively. The proposed discount function d takes into account both the above assumption and the number of scrolls required to reveal an item. Using the normalized version of (1) averaged over all users, the authors found that typical algorithms for implementing different carousel types perform differently when they are combined in a MLRI instead of being evaluated alone. Based on findings from three real-world datasets, they concluded that it is important to account for the availability of multiple collections when choosing an appropriate recommendation method. Consequently, selecting the right carousels becomes a very complex problem, which is why the authors recently proposed to use quantum computing to find a solution (Ferrari Dacrema et al., 2021). Aside from some of the aforementioned industry publications, few other authors have studied MLRI with such a holistic view. For instance, Xi et al. (2023) proposed an attentional re-ranking model that captures the user interaction with a whole page. They even went a step further by considering the special case of "F-shaped" pages, i.e., interleavings of vertical and horizontal collections, as well as the fact that users behave differently depending on the carousel type. The authors also reviewed the recent advances in page-level optimization, but these approaches are beyond the user-centered scope of this mini review.

Finally, Rahdari et al. (2022b) extended the cascade model, which describes user behavior in ranked search result lists (Craswell et al., 2008). Contrary to the above assumption, this resulted in a carousel click model that simulates user interaction under the premise that before users begin to examine the items, they explore vertically until they find a collection with a label that catches their attention. For the corresponding experiment, the authors chose labels based on genre information from the *MovieLens* dataset, i.e., only simple content-based explanations. The main finding was that the simulated users were more efficient than when scanning one-dimensional lists. Rahdari et al. (2021) also explored how to improve interactive control in MLRI by allowing users to fine-tune the importance of the topics represented by individual carousels. In a recent publication, they further demonstrated the successful use of carousels in a more practical domain, i.e., medical advice (Rahdari et al., 2022a). Without user studies, however, the empirical basis for the design of MLRI still remains weak, especially in light of other domains, where the user experience may be different depending on, e.g., item complexity and user familiarity.

2.2. User experiments on carousel recommendations

Among the publications listed in Table 1, only a few report a user experiment. Jannach et al. (2021) conducted a large online study to investigate the impact of different design alternatives (N = 775). Their exploratory study provided initial insights into the usage and assessment of carousels in the context of similar-item recommendations: Participants were slower in their decision making when they were confronted with multiple lists, but explored longer before settling on a movie. Compared to a grid without labels, the grouped organization also increased the perceived diversity and novelty of the recommended items, remarkably even with labels that did not have a meaning. With respect to labeling, the study also showed that user- and item-based carousels were preferred over references to, e.g., movie genre, director, or release date. Finally, it is worth noting that removing duplicate items did not make a difference.

Starke et al. (2022) compared a carousel-based recipe recommendation interface with a conventional vertical list and a grid (N = 150). Although the carousels had descriptive labels, they found no positive effects on choice satisfaction or difficulty compared to the grid, where the rows had no explanations and could not be scrolled horizontally. The authors noted several reasons for this finding, but it could also have been the result of the very specific task ("find the most suitable vegetarian recipe") combined with the fact that the dataset consisted only of vegetarian dishes and the labels were not very distinctive (e.g., "vegetarian recipes," "salad recipes"). However, compared to the list, both the grid and the carousels were perceived as easier to use, although it was more difficult to choose an item. Other aspects related to user experience, such as carousel length or individual decision-making traits, were not taken into account.

In another study (N = 366), however, Starke et al. (2021) examined the effects of personal characteristics and explanation styles. They found that carousels had a positive effect on choice satisfaction and perceived diversity. On the other hand, but consistent with the literature (Iyengar and Lepper, 2000), participants needed more time to make a decision than with a conventional list. While cooking experience was positively correlated with comprehensibility and satisfaction, there were no interaction effects, i.e., the MLRI had no general advantage. Moreover, no differences were found when comparing carousels with and without explanations. Apparently, labels such as "similar recipes that contain fewer calories" neither made the decision easier nor led to greater satisfaction with the chosen item. Since this contradicted earlier findings on grouped interfaces (see above), the authors concluded that it still needs to be investigated whether item details, images, or descriptive texts are more critical for making decisions than carousel labels. However, it is also important to note that the study was limited to similar-item recommendations. Given the very specific domain and the interface, which was quite different from real-world systems (few recommendations, no personalization), it is therefore difficult to generalize the results.

Only recently, Starke et al. (2023) conducted another study in their series of experiments on using carousels to promote healthy food choices (N = 164). Again, they compared a single- with a multi-list format, but also varied the personalization of the labels. While the results were consistent in terms of diversity and comprehensibility, participants were less satisfied with their choices in the multi-list condition, contrary to the findings above. Moreover, the previously observed differences in choice difficulty were not present. Regarding personalization, the authors found that labels without a focus on nutrition were preferred, e.g., "these recipes [match] your low level of cooking experience." The personalization also led to unhealthier recipe choices, possibly because participants developed negative feelings when the explanations were explicit about the relationship between personal characteristics and nutritional value. However, as acknowledged by the authors, some of the findings, including those related to the influence of health consciousness and domain knowledge, require further confirmation, especially since it was not possible to fit a structural equation model to analyze mediating effects in more depth. Besides, the relatively small recommendation sets and the fact that the crowdworkers participating in the study probably did not consume the chosen recipes may have compromised the ecological validity.

3. Summary and future research directions

The literature review has shown that in MLRI, the effects of personal characteristics and situational needs on aspects such as cognitive load and user behavior have not yet been studied to the same extent as in conventional lists. As is common in the RS field, many algorithmic advances have been proposed (see Section 2.1), but with a focus on item click data, objective metrics, and offline evaluation, partially

TABLE 1 Summary of the literature on carousel-based recommender interfaces (in chronological order).

Paper and venue/journal	Topic/contribution	Domain	Carousel types	Experiments and datasets
Wu et al. (2016) ACM RecSys conf.	Carousel and item ordering based on navigation signals	Video streaming	Various	Offline (private Netflix dataset)
McInerney et al. (2018) ACM RecSys conf.	Labeling and item ordering based on bandits	Video and music streaming	Various	Offline (private Spotify dataset), online A/B testing
Bendada et al. (2020) ACM RecSys conf.	Item ordering based on bandits, dataset, evaluation framework	Music streaming	Content-based (genres, location, mood)	Simulations (public Deezer dataset), online A/B testing
Felicioni et al. (2021) ACM IMX conf.	Offline evaluation of multi-list interfaces, evaluation metric	Video streaming	Various	Offline (MovieLens 10M dataset)
Ferrari Dacrema et al. (2021) ACM RecSys conf.	Carousel ordering using quantum computing	Video streaming	Various	Offline (MovieLens 10M, Netflix Prize dataset)
Jannach et al. (2021) ACM UMAP conf.	Study on user behavior with similar-item recommendation carousels	Video streaming	Various	Crowdsourced user study (some MovieLens dataset)
Jeunen and Goethals (2021) ACM RecSys conf.	Item ordering based on contextual bandits	Music streaming	Unspecified	Simulations (dataset from Bendada et al., 2020)
Lo et al. (2021) ACM RecSys conf.	Carousel ordering for similar-item recommendations	E-commerce	Various	Offline (private eBay dataset), online A/B testing
Rahdari et al. (2021) IntRS workshop	User control in multi-list interfaces	Education	Content-based (topics, keywords)	–
Singal et al. (2021) ACM RecSys conf.	Labeling, carousel and item ordering based on dim. reduction	Music streaming	Content-based	Offline (private Wynk Music dataset), online A/B testing
Starke et al. (2021) ACM RecSys conf.	Study on user behavior with similar-item recommendation carousels	Recipes	Content-based	Crowdsourced user study (crawled recipe dataset)
Wu et al. (2021) ACM WSDM conf.	Item ordering for 2-dim. product search based on log analysis	E-commerce	Unspecified	Offline (private Airbnb dataset)
Ferrari Dacrema et al. (2022) Frontiers in Big Data	Offline evaluation of multi-list interfaces, evaluation metric	Video streaming	Various	Offline (MovieLens 20M, Netflix Prize, ContentWise Impr. dataset)
Rahdari et al. (2022a) ACM RecSys conf.	User control in multi-list interfaces	Health-related documents	Content-based (topics)	–
Rahdari et al. (2022b) ACM HT conf.	Offline evaluation of multi-list interfaces, click model	Video streaming	Content-based (genres)	Simulations (MovieLens 100 K dataset)
Starke et al. (2022) IntRS workshop	Study on choice overload in carousels	Recipes	Content-based (categories)	Crowdsourced user study (crawled recipe dataset)
Starke et al. (2023) ACM TORS	Study on choice overload and personalization in carousels	Recipes	Content-based (categories)	Crowdsourced user studies (crawled recipe dataset)
Xi et al. (2023) ACM WSDM conf.	Carousel and item ordering based on attention networks	E-commerce	Unspecified	Offline (public Taobao dataset, crawled app store dataset)

based on assumptions that have not been validated in user experiments. In fact, there are only a few user studies available (see Section 2.2), and they do not paint a consistent picture. Instead, they have explored general design considerations in a few selected domains and with rather artificial systems[2], focusing on comparisons against conventional lists and grids, but leaving carousel-specific questions such as those raised in Section 1 and the corresponding user decision processes largely untouched.

This lack of empirical, user-centered research is particularly problematic because carousel recommendations are often personalized, but without considering the individual user experience, which also depends on aspects such as the number of carousels, their type, size, and order, as well as the selection and ranking of the items contained. Thus, we end this survey with a discussion of the directions in which future research should proceed:

- **Interface layout, carousel design, and labeling:** Specific interface aspects, such as the number and order of the collections displayed, their visible length, or the number of items they contain, still need to be investigated in user experiments with respect to their effects on decision making

2 Recommendation sets were often small, horizontal scrolling was not always possible, and the focus was often on similar-item recommendations, i.e., a reference item was visible at the top, whereas most real-world applications present self-contained carousels directly on the landing page.

and the occurrence of choice overload effects. With a better understanding of the interface layout and the design of individual carousels, it will then be possible to dynamically adjust these parameters, which in turn is a prerequisite for not only offering personalized collections, but also adapting the entire interface to the current context, i.e., improving user experience of carousel recommendations at the page level. In this regard, it may also be worth exploring how to better direct the user's attention to specific carousels, e.g., by visually highlighting relevant carousel types or adding more informative labels. Decoupled from the simple explanation styles that are currently used, but tailored more strongly to the domain, user, and current situation, this could be another important step toward reducing choice difficulty, even in an actionable way, e.g., by providing additional explanations on demand.

- **Personal characteristics and situational needs:** Any attempt to balance choice overload and the desire to explore across and within carousels will likely result in a different user experience depending on personal characteristics and situational needs: In some domains, some users may prefer a large set of diverse alternatives, while for others or in other situations, the presence of dozens of item collections may be overwhelming, possibly even leading to choice deferral (cf. Chernev et al., 2015). Thus, similar to research on one-dimensional lists, aspects such as maximization tendency (Parker et al., 2007) and decision style (Hamilton et al., 2016), but also aspects of the current context, e.g., cognitive load and domain knowledge, still need to be investigated with respect to their impact on exploration behavior (e.g., vertical and horizontal navigation depth) and selection of items from individual collections. With additional user studies, it will then be possible to draw a more consistent picture of the usage and assessment of MLRI than previous work, paving the way for more accurate user modeling, subsequent adaptation of the presentation, and ultimately better user experience.
- **Domains and datasets:** While carousels are used in almost all types of real-world applications, industry publications have only addressed e-commerce and music or video streaming. Thus, user experiments in these domains are rare, so that little is known beyond what can be inferred from clicks on the items contained in the collections. Simultaneously, few academics have conducted more user-oriented research, primarily in more serious domains, e.g., food and health (cf. Section 2.2). Given the other limitations mentioned above, it is therefore difficult to generalize their findings and to disentangle the effects of the specific use case from the influence of individual differences and aspects such as carousel type, number, and length. Moreover, existing studies did not consider item consumption, although it can strongly affect the assessment of recommendations, even in simpler domains (cf. Loepp et al., 2018). Therefore, future studies should be conducted in a wider range of domains and complemented by offline experiments and simulations. This, in turn, will require the creation of datasets that include other types of user feedback than item-related preference signals, i.e., behavioral data such as scrolling, data on the visibility and perception of carousels and items, etc.
- **Environments, devices, and modalities:** To date, MLRI have only been studied in typical web contexts, i.e., study participants had to interact with (artificial, sometimes static) web applications using a laptop or desktop computer. In practice, however, carousels are much more common on mobile devices or TVs, requiring interaction by touch or remote control. Accordingly, there is a need for studies in more naturalistic environments to better understand user behavior and decision making in relation to available carousels and interface layout. This is particularly true because the ability to satisfy diverse contexts is likely to play a much larger role in real-world applications than in the crowdsourced experiments conducted so far, where the task was predefined and focused on a single specific goal. Moreover, such studies will be useful for investigating the implementation of more explanatory labels (see above), especially if they incorporate eye-tracking analyses. Then, with richer data than item clicks, it will also be possible to validate (or reject) existing assumptions about user behavior and to obtain more comprehensive models of user interaction, which can subsequently be used to further improve the user experience of MLRI.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

We acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Beel, J., and Dixon, H. M. (2021). "The 'unreasonable' effectiveness of graphical user interfaces for recommender systems," in *UMAP '21: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (New York, NY: ACM), 22–28. doi: 10.1145/3450614.3461682

Bendada, W., Salha, G., and Bontempelli, T. (2020). "Carousel personalization in music streaming apps with contextual bandits," in *RecSys '20: Proceedings of the 14th ACM Conference on Recommender Systems* (New York, NY: ACM), 420–425. doi: 10.1145/3383313.3412217

Bollen, D., Knijnenburg, B. P., Willemsen, M. C., and Graus, M. P. (2010). "Understanding choice overload in recommender systems," in *RecSys '10: Proceedings of the 4th ACM Conference on Recommender Systems* (New York, NY: ACM), 63–70. doi: 10.1145/1864708.1864724

Chen, L., and Pu, P. (2010). "Eye-tracking study of user behavior in recommender interfaces," in *UMAP '10: Proceedings of the 18th International Conference on User Modeling, Adaptation and Personalization* (Berlin: Springer), 375–380. doi: 10.1007/978-3-642-13470-8_35

Chen, L., and Pu, P. (2012a). Critiquing-based recommenders: survey and emerging trends. *User Model. User Adapt. Interact.* 22, 125–150. doi: 10.1007/s11257-011-9108-6

Chen, L., and Pu, P. (2012b). Experiments on user experiences with recommender interfaces. *Behav. Inform. Technol.* 33, 372–394. doi: 10.1080/0144929X.2012.719034

Chernev, A., Böckenholt, U., and Goodman, J. (2015). Choice overload: a conceptual review and meta-analysis. *J. Cons. Psychol.* 25, 333–358. doi: 10.1016/j.jcps.2014.08.002

Craswell, N., Zoeter, O., Taylor, M., and Ramsey, B. (2008). "An experimental comparison of click position-bias models," in *WSDM '08: Proceedings of the 1st ACM International Conference on Web Search and Data Mining* (New York, NY: ACM), 87–94. doi: 10.1145/1341531.1341545

Felicioni, N., Ferrari Dacrema, M., and Cremonesi, P. (2021). "Measuring the user satisfaction in a recommendation interface with multiple carousels," in *IMX '21: Proceedings of the 2nd ACM International Conference on Interactive Media Experiences* (New York, NY: ACM), 212–217. doi: 10.1145/3452918.3465493

Ferrari Dacrema, M., Felicioni, N., and Cremonesi, P. (2021). "Optimizing the selection of recommendation carousels with quantum computing," in *RecSys '21: Proceedings of the 15th ACM Conference on Recommender Systems* (New York, NY: ACM), 691–696. doi: 10.1145/3460231.3478853

Ferrari Dacrema, M., Felicioni, N., and Cremonesi, P. (2022). Offline evaluation of recommender systems in a user interface with multiple carousels. *Front. Big Data* 5:910030. doi: 10.3389/fdata.2022.910030

Gomez-Uribe, C. A., and Hunt, N. (2015). The Netflix recommender system: algorithms, business value, and innovation. *ACM Trans. Manage. Inform. Syst.* 6, 13:1–13:19. doi: 10.1145/2843948

Guntuku, S. C., Roy, S., Lin, W., Ng, K., Keong, N. W., and Jakhetiya, V. (2016). "Personalizing user interfaces for improving quality of experience in VoD recommender systems," in *QoMEX '16: Proceedings of the 8th International Conference on Quality of Multimedia Experience* (Washington, DC). doi: 10.1109/QoMEX.2016.7498940

Hamilton, K., Shih, S.-I., and Mohammed, S. (2016). The development and validation of the rational and intuitive decision styles scale. *J. Pers. Assess.* 98, 523–535. doi: 10.1080/00223891.2015.1132426

He, C., Parra, D., and Verbert, K. (2016). Interactive recommender systems: a survey of the state of the art and future research challenges and opportunities. *Expert Syst. Appl.* 56, 9–27. doi: 10.1016/j.eswa.2016.02.013

Iyengar, S. S., and Lepper, M. R. (2000). When choice is demotivating: can one desire too much of a good thing? *J. Pers. Soc. Psychol.* 79, 995–1006. doi: 10.1037/0022-3514.79.6.995

Jannach, D., Jesse, M., Jugovac, M., and Trattner, C. (2021). "Exploring multi-list user interfaces for similar-item recommendations," in *UMAP '21: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (New York, NY: ACM), 224–228. doi: 10.1145/3450613.3456809

Jeunen, O., and Goethals, B. (2021). "Top-K contextual bandits with equity of exposure," in *RecSys '21: Proceedings of the 15th ACM Conference on Recommender Systems* (New York, NY: ACM), 310–320. doi: 10.1145/3460231.3474248

Jin, Y., Tintarev, N., Htun, N. N., and Verbert, K. (2020). Effects of personal characteristics in control-oriented user interfaces for music recommender systems. *User Model. User Adapt. Interact.* 30, 199–249. doi: 10.1007/s11257-019-09247-2

Jugovac, M., and Jannach, D. (2017). Interacting with recommenders-overview and research directions. *ACM Trans. Interact. Intell. Syst.* 7, 10:1–10:46. doi: 10.1145/3001837

Kammerer, Y., and Gerjets, P. (2010). "How the interface design influences users' spontaneous trustworthiness evaluations of web search results: comparing a list and a grid interface,"' in *ETRA '10: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (New York, NY: ACM), 299–306. doi: 10.1145/1743666.1743736

Kleemann, T., Wagner, M., Loepp, B., and Ziegler, J. (2021). "Modeling user interaction at the convergence of filtering mechanisms, recommender algorithms and advisory components," in *Mensch & Computer 2021-Tagungsband* (New York, NY: ACM), 531–543. doi: 10.1145/3473856.3473859

Knijnenburg, B. P., and Willemsen, M. C. (2015). "Evaluating recommender systems with user experiments," in *Recommender Systems Handbook* (Boston, MA: Springer US), 309–352. doi: 10.1007/978-1-4899-7637-6_9

Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., and Getoor, L. (2019). "Personalized explanations for hybrid recommender systems," in *IUI '19: Proceedings of the 24th International Conference on Intelligent User Interfaces* (New York, NY: ACM), 379–390. doi: 10.1145/3301275.3302306

Kunkel, J., and Ziegler, J. (2023). A comparative study of item space visualizations for recommender systems. *Int. J. Hum. Comput. Stud.* 172:102987. doi: 10.1016/j.ijhcs.2022.102987

Lo, C., Yu, H., Yin, X., Shetty, K., He, C., Hu, K., et al. (2021). "Page-level optimization of e-commerce item recommendations," in *RecSys '21: Proceedings of the 15th ACM Conference on Recommender Systems* (New York, NY: ACM), 495–504. doi: 10.1145/3460231.3474242

Loepp, B., Donkers, T., Kleemann, T., and Ziegler, J. (2018). "Impact of item consumption on assessment of recommendations in user studies," in *RecSys '18: Proceedings of the 12th ACM Conference on Recommender Systems* (New York, NY: ACM), 49–53. doi: 10.1145/3240323.3240375

Loepp, B., Donkers, T., Kleemann, T., and Ziegler, J. (2019). Interactive recommending with tag-enhanced matrix factorization (TagMF). *Int. J. Hum. Comput. Stud.* 121, 21–41. doi: 10.1016/j.ijhcs.2018.05.002

Lousame, F. P. P., and Sánchez, E. (2009). "View-based recommender systems," in *RecSys '09: Proceedings of the 3rd ACM Conference on Recommender Systems* (New York, NY: ACM), 389–392. doi: 10.1145/1639714.1639795

McInerney, J., Lacker, B., Hansen, S., Higley, K., Bouchard, H., Gruson, A., et al. (2018). "Explore, exploit, and explain: personalizing explainable recommendations with bandits," in *RecSys '18: Proceedings of the 12th ACM Conference on Recommender Systems* (New York, NY: ACM), 31–39. doi: 10.1145/3240323.3240354

Millecamp, M., Htun, N. N., Conati, C., and Verbert, K. (2019). "To explain or not to explain: the effects of personal characteristics when explaining music recommendations," in *IUI '19: Proceedings of the 24th International Conference on Intelligent User Interfaces* (New York, NY: ACM), 397–407. doi: 10.1145/3301275.3302313

Nanou, T., Lekakos, G., and Fouskas, K. (2010). The effects of recommendations' presentation on persuasion and satisfaction in a movie recommender system. *Multim. Syst.* 16, 219–230. doi: 10.1007/s00530-010-0190-0

Nazari, Z., Chandar, P., Fazelnia, G., Edwards, C. M., Carterette, B., and Lalmas, M. (2022). "Choice of implicit signal matters: Accounting for user aspirations in podcast recommendations," in *WWW '22: Proceedings of the 31st ACM Web Conference* (New York, NY: ACM), 2433–2441. doi: 10.1145/3485447.3512115

Parker, A. M., De Bruin, W. B., and Fischhoff, B. (2007). Maximizers vs. satisficers: decision-making styles, competence, and outcomes. *Judgm. Decis. Mak.* 2:342. doi: 10.1017/S1930297500000486

Rahdari, B., Brusilovsky, P., He, D., Thaker, K. M., Luo, Z., and Lee, Y. J. (2022a). "HELPeR: an interactive recommender system for ovarian cancer patients and caregivers," in *RecSys '22: Proceedings of the 16th ACM Conference on Recommender Systems* (New York, NY: ACM), 644–647. doi: 10.1145/3523227.3551471

Rahdari, B., Brusilovsky, P., and Sabet, A. J. (2021). "Controlling personalized recommendations in two dimensions with a carousel-based interface," in *IntRS '21: Proceedings of the 8th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems*, 112–122. Available online at: https://ceur-ws.org/Vol-2948/.

Rahdari, B., Kveton, B., and Brusilovsky, P. (2022b). "The magic of carousels: single vs. multi-list recommender systems," in *HT '22: Proceedings of the 33rd ACM Conference on Hypertext and Social Media* (New York, NY: ACM), 166–174. doi: 10.1145/3511095.3531278

Scheibehenne, B., Greifeneder, R., and Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *J. Cons. Res.* 37, 409–425. doi: 10.1086/651235

Singal, S., Singh, P., and Dahiya, M. (2021). "Automatic collection creation and recommendation," in *RecSys '21: Proceedings of the 15th ACM Conference on Recommender Systems* (New York, NY: ACM), 633–638. doi: 10.1145/3460231.3478865

Starke, A., Sedkowska, J., Chouhan, M., and Ferwerda, B. (2022). "Examining choice overload across single-list and multi-list user interfaces," in *IntRS '22: Proceedings of the 8th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems*.

Starke, A. D., Asotic, E., and Trattner, C. (2021). "'Serving each user': Supporting different eating goals through a multi-list recommender interface," in *RecSys '21: Proceedings of the 15th ACM Conference on Recommender Systems* (New York, NY: ACM), 124–132. doi: 10.1145/3460231.3474232

Starke, A. D., Asotic, E., Trattner, C., and Van Loo, E. J. (2023). Examining the user evaluation of multi-list recommender interfaces in the context of healthy recipe choices. *ACM Trans. Recomm. Syst*. doi: 10.1145/3581930

Wu, C.-Y., Alvino, C. V., Smola, A. J., and Basilico, J. (2016). "Using navigation to improve recommendations in real-time," in *RecSys '16: Proceedings of the 10th ACM Conference on Recommender Systems* (New York, NY: ACM), 341–348. doi: 10.1145/2959100.2959174

Wu, L., Grbovic, M., and Li, J. (2021). "Toward user engagement optimization in 2d presentation," in *WSDM '21: Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (New York, NY: ACM), 1047–1055. doi: 10.1145/3437963.3441749

Xi, Y., Lin, J., Liu, W., Dai, X., Zhang, W., Zhang, R., et al. (2023). "A bird's-eye view of reranking: from list level to page level," in *WSDM '23: Proceedings of the 16th ACM International Conference on Web Search and Data Mining* (New York, NY: ACM), 1075–1083. doi: 10.1145/3539597.3570399

frontiers | Frontiers in Big Data

TYPE Review
PUBLISHED 06 October 2023
DOI 10.3389/fdata.2023.1245198

Check for updates

OPEN ACCESS

EDITED BY
Deqing Yang,
Fudan University, China

REVIEWED BY
Aonghus Lawlor,
University College Dublin, Ireland
Shaohua Tao,
Xuchang University, China

*CORRESPONDENCE
Markus Schedl
✉ markus.schedl@jku.at

RECEIVED 23 June 2023
ACCEPTED 18 September 2023
PUBLISHED 06 October 2023

CITATION
Kumar D, Grosz T, Rekabsaz N, Greif E and Schedl M (2023) Fairness of recommender systems in the recruitment domain: an analysis from technical and legal perspectives. *Front. Big Data* 6:1245198.
doi: 10.3389/fdata.2023.1245198

COPYRIGHT
© 2023 Kumar, Grosz, Rekabsaz, Greif and Schedl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Fairness of recommender systems in the recruitment domain: an analysis from technical and legal perspectives

Deepak Kumar[1], Tessa Grosz[2], Navid Rekabsaz[1,3], Elisabeth Greif[2] and Markus Schedl[1,3]*

[1]Multimedia Mining and Search Group, Institute of Computational Perception, Johannes Kepler University Linz, Linz, Austria, [2]Institute for Legal Gender Studies, Johannes Kepler University Linz, Linz, Austria, [3]Human-centered AI Group, AI Lab, Linz Institute of Technology, Linz, Austria

Recommender systems (RSs) have become an integral part of the hiring process, be it via job advertisement ranking systems (job recommenders) for the potential employee or candidate ranking systems (candidate recommenders) for the employer. As seen in other domains, RSs are prone to harmful biases, unfair algorithmic behavior, and even discrimination in a legal sense. Some cases, such as salary equity in regards to gender (gender pay gap), stereotypical job perceptions along gendered lines, or biases toward other subgroups sharing specific characteristics in candidate recommenders, can have profound ethical and legal implications. In this survey, we discuss the current state of fairness research considering the fairness definitions (e.g., demographic parity and equal opportunity) used in recruitment-related RSs (RRSs). We investigate from a technical perspective the approaches to improve fairness, like synthetic data generation, adversarial training, protected subgroup distributional constraints, and *post-hoc* re-ranking. Thereafter, from a legal perspective, we contrast the fairness definitions and the effects of the aforementioned approaches with existing EU and US law requirements for employment and occupation, and second, we ascertain whether and to what extent EU and US law permits such approaches to improve fairness. We finally discuss the advances that RSs have made in terms of fairness in the recruitment domain, compare them with those made in other domains, and outline existing open challenges.

KEYWORDS

recommender system, recruitment, job recommendation, candidate recommendation, fairness, discrimination, law

1. Introduction

Recommender systems (RSs) in the recruitment domain are usable by both job seekers (job recommenders) and candidate seekers (candidate recommenders). An early application of the recruitment-related RSs (RRSs) is CASPER (Case-Based Profiling for Electronic Recruitment) (Rafter et al., 2000), an automated collaborative filtering-based personalized case retrieval system. Most modern RRSs, such as LinkedIn, have diversified their approaches and use a variety of other methods such as exploiting textual data available in the recruitment domain, or social network knowledge (Fawaz, 2019; Geyik et al., 2019). In terms of algorithms, we have come a long way from linguistics-based systems (Vega, 1990) to the current RSs which are based on deep neural networks, collaborative filtering, content-based, and knowledge-based techniques (Gutiérrez et al., 2019; Bian et al., 2020; Gugnani and Misra, 2020; Lacic et al., 2020).

However, the prevalent use of RSs has also highlighted the possibility of biased outcomes in the recruitment domain. For instance, the gender stereotypes pertaining to particular professions are observed in the current workforce (Wilson et al., 2014; Smith et al., 2021). These stereotypes can further find their way into RRSs in the form of algorithmic bias (Tang et al., 2017; Ali et al., 2019; Raghavan et al., 2020). Algorithmic bias can cause discrimination in the exposure of the job advertisement or the algorithmic hiring itself. Facebook's advertisement delivery system, for example, suffered from algorithmic gender bias while showing job advertisements (Ali et al., 2019). The Amazon hiring algorithm, infamously favoring male over female job applicants, is another real-world example (Dastin, 2018). The historical data provided to the algorithm suggested that male applicants were preferred because previously more men than women had been hired. Such behaviors prompted the industry to adopt bias mitigation in RRSs (Raghavan et al., 2020).

Such algorithmic biases can have legal consequences. Hiring decisions, whether algorithmically assisted through RRSs or solely taken by the employer, are part of the employment process and as such do not operate in a legal vacuum: Non-discrimination law plays an essential role in safeguarding from discrimination in the recruitment domain and should not be overlooked by RRS researchers. Given this context, the objective of this survey is threefold:

- to examine recent studies focusing on fairness in RRSs,
- to emphasize the significant disparities between the research conducted in the fields of computer science (CS) and law concerning this topic, and
- to identify the challenges that exist within fairness research for RSs employed in the recruitment domain.

1.1. Related surveys

There exist multiple surveys on the fairness of RSs (Zehlike et al., 2022; Deldjoo et al., 2023; Wang et al., 2023). While these surveys partly covers work on RRSs, they are not tailored to this domain, nor do they connect the discussion of fairness to the legal aspects of this domain. These two points are crucial, particularly due to the essential differences of RRSs from those used in other domains (e.g., video, music, or e-commerce): First, the decision of the system may have significant impacts on the end-users in terms of fairness and distribution of resources and can have serious legal implications. Second, the recruitment domain heavily relies on textual data of partly personal and sensitive nature, for instance, the resumes, and job posts which can contain sensitive information about candidates and employers, respectively. There also exist several surveys on RRSs, e.g., Dhameliya and Desai (2019), de Ruijt and Bhulai (2021), Freire and de Castro (2021), and Thali et al. (2023), but only de Ruijt and Bhulai (2021) include a dedicated section covering fairness aspects. However, de Ruijt and Bhulai (2021) provide a broad scope on the topic, and only present high-level insights on fairness aspects. In particular, we are not aware of any survey in this domain that draws the connection between algorithmic aspects of RRSs and the intertwined legal facets. To fill these gaps in existing surveys on the fairness of RSs and RRSs, the survey at hand calls for attention to the research specifically addressing the fairness of RSs in the recruitment domain, with a multidisciplinary analysis from both technical and legal perspectives.

TABLE 1 List of abbreviations used in the article. For abbreviations of fairness metrics, please refer to Table 3.

AI	Artificial Intelligence
CBF	Content-based Filtering
CF	Collaborative Filtering
CJEU	Court of Justice of the European Union
CS	Computer Science
DP	Demographic Parity
EO	Equal Opportunity
IF	Individual Fairness
KB	Knowledge-Based
PF	Proportional Fairness
RRS	Recruitment-related Recommender System
RS	Recommender System

1.2. Literature search

To identify relevant literature, we conducted a series of searches on DBLP[1] with the keywords "Job"/"Candidate" in conjunction with "Fair"/"Bias" to create a candidate list of publications for this survey. Subsequently, we selected a subset of this list based on their relevance to RSs in the recruitment domain after a careful manual inspection of each paper. We further enriched the list with articles from the workshop series on Recommender Systems for Human Resources (RecSys in HR) and with further recent works spotted by studying the references in the collected papers. The surveyed literature covers the work published until May 2023.

1.3. Outline

The survey is structured as follows: Section 2 provides an overview of the current research on fairness in RRSs. In Section 3, we focus on the significant legal aspects that pertain to RSs in the recruitment domain, aiming to bridge the gap between the legal and technical dimensions of ensuring fairness in candidate recommenders, and to highlight the shortcomings of the existing approaches for fairness in RS. Finally, Section 4 discusses the open challenges and future directions related to fairness research in RRSs. For easy reference, Table 1 contains a list of abbreviations used throughout the article.

1 https://dblp.org/ (accessed May 2023).

TABLE 2 Overview and categorization of the research works surveyed.

Article	RS		Dataset(s)	Fairness measurement		Debiasing strategy	Sensitive attribute(s)
	Type	Algorithm		Definition(s)	Metric(s)		
Li et al. (2023)	Job	CF	Private	DP	UGF	Post	Gender (G)
Rus et al. (2022)	Job	CBF	Private	EO, DP	TPRP, SAT	Pre, In	G
Ntioudis et al. (2022)	Job	KB	Synthetic, private	—	—	Pre	Migrant
Zhang (2021)	Job	—	Synthetic	DP	SDR, LDR	—	G
Shishehchi and Banihashem (2019)	Job	KB	Private	—	UAT	Pre	Disability
Tang et al. (2017)	Job	—	Scraped	—	MWU	—	G
Scher et al. (2023)	Candidate	—	Synthetic	EO, DP	BGSD, CF, TNR	Post	—
Jourdan et al. (2023)	Candidate	CBF	BIOS	EO	TPRP	Pre, In	G
Delecraz et al. (2022)	Candidate	—	Private	EO	TPRP, DI, SP	—	Multi
Markert et al. (2022)	Candidate	CBF	Synthetic	IF	IFC	—	G, Marital status
Bei et al. (2020)	Candidate	Other	Synthetic	PF	Violations	Post	G, Region
Arafan et al. (2022)	Candidate	—	Synthetic, Private	DP	NDKL	Pre, Post	G
Tran et al. (2022)	Candidate	KB	ACI, TO, HR	—	—	Post	Disability
Syed and Shivendu (2022)	Candidate	—	—	EO, DP	TPRP	Pre	—
Burke et al. (2021)	Candidate	Other	SIOP2021	EO	AIR	Post	—
Tran et al. (2021)	Candidate	KB	Synthetic, Private	—	—	Post	Disability
Wilson et al. (2021)	Candidate	—	Private	EO, DP	AIR, MWU, KWH	—	G, Ethnicity
Elbassuoni et al. (2020)	Candidate	—	Synthetic, Private	DP	EMD	—	G, Ethnicity
Elbassuoni et al. (2019)	Candidate	—	Private	DP	EMD	—	Multi
Geyik et al. (2019)	Candidate	—	Synthetic, Private	EO, DP	NDKL, MS, II, IC	Post	G, Age
Chen et al. (2018)	Candidate	—	Scraped	DP, IF	MWU, TDRC, EC	—	G
Amer-Yahia et al. (2020)	Both	—	Synthetic, Private	DP	KT, EMD, Exposure	—	G, Ethnicity, Nationality

In RS Type, "Both" is used when both job recommender and candidate recommender are used, and for Sensitive Attribute(s), "Multi" is used when there are four or more attributes (e.g., Gender, Nationality, Age, Ethnicity, Language). The symbol "—" is used when the column is not applicable or cannot be inferred from the paper.

2. Current state of research

As a core reference for interested readers, we provide in Table 2 an overview and categorization of existing research that addresses fairness in RRSs. The individual works are classified based on the datasets utilized for the experiments, the aspired fairness definitions, the metrics employed to assess fairness, the stage in the pipeline where debiasing strategies are implemented to achieve fairness, and the sensitive attributes explored during the experiments.

This section analyzes the current state of fairness research in RRSs. Our examination of the existing literature is structured around various aspects, including the *recommendation algorithm* underlying the RRSs (Section 2.1), the *datasets* used in the reviewed works (Section 2.2), the *definition of fairness* (Section 2.3), the *metrics* used to evaluate fairness or unfairness in the datasets and created recommendations (Section 2.4), *evidence of bias* within these algorithms (Section 2.5), and the different *approaches* explored to attain fairness (Section 2.6).

In the article at hand, we adopt the terminology common in RSs research, i.e., we refer to "users" (U) and "items" (I), the former representing the entity for which recommendations are made (input), the latter the recommendations themselves (output). Hence, for a job recommender system, the user is the candidate looking for a job, and the items are job posts. In contrast, for a candidate recommender, the user is the job post and the items are candidates.

2.1. Recommendation algorithms in the recruitment domain

RRSs can be classified into four categories: collaborative filtering (CF), content-based filtering (CBF), knowledge-based

(KB) recommenders, and hybrid recommenders. In research on fairness of RRSs, company platforms are mostly investigated, e.g., Indeed, Monster, and CareerBuilder[2] in Chen et al. (2018); LinkedIn in Tang et al. (2017) and Geyik et al. (2019); TaskRabbit and Google Job Search[3] in Amer-Yahia et al. (2020) and Elbassuoni et al. (2020). These works are about investigating fairness of these popular platforms rather than improving it.

In contrast to the proprietary and commonly non-disclosed recommendation algorithms used by the aforementioned services, RRS approaches found in the surveyed literature can be categorized into the following:

- Collaborative Filtering (CF): CF algorithms employed in RRSs are all based on the similarity between users, calculated through the users' interactions with the items. In the RRSs literature, users are always job seekers, i.e., no CF approaches are used for candidate recommenders, only job recommenders. The adopted algorithms include probabilistic matrix factorization (Salakhutdinov and Mnih, 2007), neural matrix factorization (He et al., 2017), session-based model STAMP (Short-Term Attention/Memory Priority) (Liu et al., 2018), and matrix factorization with global bias terms (Koren et al., 2009).
- Content-based Filtering (CBF): CBF algorithms are based on the similarity between the items and the users, most commonly implemented as direct matching of users and items through text-based similarity. In the surveyed RRSs literature, users can be job posts (candidate recommenders) or candidates (job recommenders). Rus et al. (2022) and Jourdan et al. (2023) implement RoBERTa (Liu et al., 2019) and BERT (Devlin et al., 2019) language models, respectively, to learn candidate-job similarity. Markert et al. (2022) use a custom regression model to learn the candidate-job similarity. In these works, both language model and regression model are neural network-based models.
- Knowledge-based (KB): The algorithms in this category utilize domain-specific knowledge to create ontology and compute similarity between user-item pairs. Existing works that fall in this category use directed acyclic causal graphs (Tran et al., 2021, 2022) and knowledge representation using domain ontology (Shishehchi and Banihashem, 2019; Ntioudis et al., 2022).
- Hybrid: Hybrid approaches utilize combinations of the previous approaches (Luo et al., 2019). No such hybrid could be found in the conducted literature search.
- Others: We group here the studies that do not use the mentioned recommendation approaches. In particular, Burke et al. (2021) implements spatial search-based candidate selection, and Bei et al. (2020) deploys integer linear programming-based candidate selection.

2 https://www.indeed.jobs/, https://www.monster.com/, https://www.careerbuilder.co.uk/ (accessed May 2023).

3 https://www.taskrabbit.com/, https://jobs.google.com/about/ (accessed May 2023).

2.2. Datasets

In Table 2, we divide the datasets into four categories: dataset that cannot be accessed or recreated (denoted *Private* in the table), scraping details are given (*Scraped*), procedure to create the artificial dataset is given (*Synthetic*), and name of the dataset if it is public. The datasets used in the reviewed works are primarily private or synthetically created. Table 2 also reveals the surprising absence of popular public job recommender datasets, such as the datasets used in the ACM Recommender Systems Challenge 2016 and 2017 (Abel et al., 2016, 2017) and the Career Builder 2012 dataset.[4] Public datasets are used only for candidate recommenders.

- *BIOS (De-Arteaga et al., 2019):* This dataset has been created by scrapping biographies using Common Crawl.[5] The dataset contains biographies of individuals with the attributes current job and gender. The dataset proposed by the authors contains 397,340 biographies with 28 different occupations.
- *Adult Census Income (ACI) (Becker and Kohavi, 1996):* The ACI dataset has been extracted from the 1994 USA census database and includes 48,842 individuals' records. For each individual, it contains features such as occupation, age, education, marital status, salary, race, and gender.
- *TO*[6]: The dataset contains 1,129 Russian workers' income and 16 features describing them, such as gender, age, profession, experience, industry, employee turnover, supervisor, supervisor's gender, and recruitment route.
- *HR*[7]: This dataset contains 15,000 individuals' retention records with the features satisfaction level, last evaluation, average monthly hours, work accident, salary, and time spent in company.
- *SIOP2021 (Koenig and Thompson, 2021):* This dataset was introduced in the machine learning challenge of the Annual Conference of the Society for Industrial and Organizational Psychology 2021. The dataset contains the three attributes performance, turnover data, and the protected group membership[8] for 7,890 respondents.

2.3. Fairness definitions for recruitment-related recommender systems

We review different fairness definitions in RRSs in this section. Overall, as shown in Table 2, one of the most used fairness definitions in the context of RRSs is demographic parity (DP). Also, the fairness definition for job recommenders is restricted to DP in the identified literature. Overall, DP is the

4 https://www.kaggle.com/competitions/job-recommendation/overview (accessed June 2023).

5 https://commoncrawl.org/ (accessed June 2023).

6 https://www.kaggle.com/datasets/davinwijaya/employee-turnover (accessed June 2023).

7 https://www.kaggle.com/datasets/liujiaqi/hr-comma-sepcsv (accessed June 2023).

8 The actual protected group is not mentioned by the creators.

most often used fairness definition for both job and candidate recommenders. We will see in the definition of DP that it does not consider the quality of recommendation (i.e., whether relevant items are ranked high or low is ignored. This results in an easier adaptation of this fairness definition from classification to recommendation tasks for both candidate recommenders and job recommenders.

Most fairness definitions mentioned in the literature are adapted from the binary classification setting to the RS setting. The core fairness definitions from the classification are listed below:

- *Demographic parity (DP):* A binary predictor $\hat{Y}$ is said to satisfy demographic parity with respect to protected attribute $A \in \{a_1, ..., a_l\}$ that can take l values if $\hat{Y}$ is independent of A (Dwork et al., 2012).

$$P(\hat{Y}|A = a_1) = ... = P(\hat{Y}|A = a_l)$$

- *Equal opportunity (EO):* A binary predictor $\hat{Y}$ is said to satisfy equal opportunity with respect to a protected attribute $A \in \{a_1, ..., a_l\}$ that can take l values and ground truth Y if they are independent conditional on the ground truth outcome being favorable (Hardt et al., 2016).

$$P(\hat{Y} = 1|A = a_1, Y = 1) = ... = P(\hat{Y} = 1|A = a_l, Y = 1)$$

- *Individual fairness (IF):* A predictor $\hat{Y}$ is said to satisfy individual fairness if for data-point x from dataset D for all x' that are similar to x (i.e., all their attributes are the same except for the sensitive attribute) the predictor predicts the same class (Ruoss et al., 2020).

$$x \sim D, \forall x' \in \mathbb{R} : \phi(x, x') \implies \mu(\hat{Y}_x, \hat{Y}_{x'})$$

where $\phi(x, x') = 1$ iff x and x' are similar
and $\mu(\hat{Y}_x, \hat{Y}_{x'}) = 1 \iff \hat{Y}_x = \hat{Y}_{x'}$

As we explain in the following, the fairness definitions for candidate recommenders are strongly aligned with classification scenarios, i.e., whether an item (candidate) occurs in the recommendation list or not. At the same time, fairness definitions for job recommenders require further modification. The fairness definitions in RRSs are always from the candidate's perspective. In the case of job recommenders, the involved variables are sensitive attributes A_u (e.g., gender or ethnicity) of the input u (i.e., candidate), and the recommended list Q_u (i.e., list of job posts). For candidate recommenders, the sensitive attribute A_i belongs to the item i in the recommended list Q_u (i.e., list of candidates). Furthermore, defining fairness for job recommenders requires the function $\mathcal{F}$ defined over the recommended list Q_u (list of job posts), which measures the quality of recommendation (e.g., through precision or recall metrics) or some other property of Q_u like the average salary of jobs in Q_u. Based on these definitions, in the following, we review the fairness definitions for RRSs, adapted from classification tasks:

- *DP for job recommender (Li et al., 2023)*: A job recommender satisfies DP for sensitive attribute gender $A_u \in \{male, female, non-binary\}$ if some measure, expressed as a function $\mathcal{F}$ defined over the recommendation list of jobs, is independent of the value of A_u, i.e.,

$$P(\mathcal{F}(\mathbf{Q}_u)|A_u = male) = P(\mathcal{F}(\mathbf{Q}_u)|A_u = female) \\ = P(\mathcal{F}(\mathbf{Q}_u)|A_u = non-binary)$$

Example: The average salary of jobs recommended have the same distribution for the male, female, and non-binary candidate groups.

- *DP for candidate recommender (Geyik et al., 2019)*: A candidate recommender satisfies DP for attribute tuple $A_i =< gender_i, age_i >$, if the existence or absence of any candidate i in the recommended list Q_u for job ad u is independent of their attributes A_i.

$$P(i \in Q_u|A_i =< gender_i, age_i >) = P(j \in Q_u|A_j =< gender_j, age_j >)$$
$$P(i \notin Q_u|A_i =< gender_i, age_i >) = P(j \notin Q_u|A_j =< gender_j, age_j >)$$

Example: A young male has the same probability as an old female of being included in the recommendation list for the job.

- *EO for candidate recommender (Geyik et al., 2019)*: A candidate recommender satisfies EO with respect to protected attribute tuple $A_i =< gender_i, age_i >$ if candidate i's existence in the recommended list Q_u with respect to protected attributes A_i is conditionally independent of the candidate being qualified [i.e. $\rho(i, u) = 1$] for the job u.

$$P(i \in Q_u|A_i =< gender_i, age_i >, \rho(i, u) = 1) \\ = P(j \in Q_u|A_j =< gender_j, age_j >, \rho(i, u) = 1)$$

Example: A young male who satisfies the requirements for the job has the same probability of being selected for the job as an old female that satisfies the same requirements.

- *IF for candidate recommender (Markert et al., 2022)*: A candidate recommender satisfies IF if for an item i (i.e., candidate) in the recommended list Q_u (for a job ad u) all the candidates i' similar to i (i.e., their attributes are the same except for the sensitive attribute) in the recommended lists have nearby positions in the ranking.

$$i \in Q_u, \forall i' \in Q_u : \phi(i, i') \implies \mu(Q_u, i, i')$$

where $\phi(i, i') = 1$ iff i and i' are similar
and $\mu(Q_u, i, i') = 1 \iff pos(i, Q_u) \approx pos(i', Q_u)$
Example: If two candidates of different gender in the item set have the same attributes, they should be ranked at nearby positions in the recommendation list.

In contrast to the mentioned definitions, *Proportional Fairness (Bei et al., 2020)* is not adapted from classification and is directly defined for RSs.

- *Proportional Fairness (PF)* (Bei et al., 2020): For a candidate recommender to satisfy PF, the selected set of candidates S and candidate attribute set $A = \{a_1, ..., a_l\}$, at least fraction α_j and at most fraction β_j of S have attribute a_j.
Example: Five candidates are selected for a given job by the RRS. The attributes of these people are gender (male vs. female) and region (Europe vs. Africa). Then, for proportional fairness over the attributes male, female, Europe, and Africa under the constraints (least fraction, most fraction), respectively, (0.4, 0.6), (0.4, 0.6), (0.2, 0.4), and (0.2, 1.0), it is required that the number of people out of 5 with attribute male, female, Europe, and Africa are {2, 3},{2, 3},{1, 2}, and {1, 2, 3, 4, 5}, respectively.

2.4. Fairness and unfairness metrics

Formalizing the fairness definitions introduced above, the RRSs literature has proposed various metrics to quantify the degree to which fairness is achieved by a given RRS. In this section, we therefore present the fairness metrics used in the surveyed literature. As we could already see in Table 2, they are very diverse. The only recurring metrics are TRPR, EMD, NDKL, MWU, and AIR. The frequency of each metric's use in literature can be seen in Table 3, along with a categorization of the metrics with respect to the targetted fairness definitions. Below we introduce the common metrics and the metrics we refer to later in the manuscript. For the remaining ones, we refer the reader to the corresponding references provided in Table 2.

The metrics in the literature addressing DP consider the distribution of sensitive attributes over recommended lists, as shown for here:

- *User Group Fairness (UGF)* (Li et al., 2023): Given metric $\mathcal{F}$ (e.g., average salary) over a recommended list of jobs Q_u for the user u (i.e. candidate) and the set of all candidates U partitioned in two mutually exclusive sets U_{a_1}, U_{a_2} based on protected attribute $A \in \{a_1, a_2\}$, UGF is defined as:

$$UGF(U_{a_1}, U_{a_2}, Q) = \left| \frac{1}{|U_{a_1}|} \sum_{u \in U_{a_1}} \mathcal{F}(\mathbf{Q}_u) - \frac{1}{|U_{a_2}|} \sum_{u \in U_{a_2}} \mathcal{F}(\mathbf{Q}_u) \right|$$

where $Q = \{Q_u \forall u \in U\}$ is the set of all recommendation lists.

- *Set Difference Rate (SDR)* (Zhang, 2021): SDR is defined to measure proportion of attribute-specific jobs. SDR between set of items (i.e., jobs) I_{a_1} and I_{a_2} recommended only to users (i.e., candidates) with sensitive attribute $A = a_1$ and $A = a_2$ (where $A \in \{a_1, a_2\}$) respectively and the set of all jobs I.

$$SDR(I_{a_1}, I_{a_2}) = \frac{|I_{a_1}| + |I_{a_2}|}{|I|}$$

- *List Difference Rate (LDR)* (Zhang, 2021): LDR is defined to measure differences in ranking due to the binary protected attribute $A \in \{a_1, a_2\}$ of a user. LDR for a pair of recommendation lists Q_u and $Q_{\hat{u}}$ of a candidate u and its counterfactual $\hat{u}$ (created by changing u's protected attribute), respectively, is

$$LDR(Q_u, Q_{\hat{u}}) = \frac{\sum_{j=1}^{|Q_u|} \gamma\left(Q_u, Q_{\hat{u}}, j\right)}{|Q_u|}$$

where $\gamma(Q_u, Q_{\hat{u}}, j) = 1$ if the j^{th} job ad in Q_u and $Q_{\hat{u}}$ is the same.

- *Earth Mover's Distance (EMD)* (Pele and Werman, 2009): EMD can be defined for sensitive attribute $A \in \{a_1, ..., a_l\}$ that can take l values by dividing the ranking Q_u in two groups, group G with $A = a_1$ and group $\bar{G}$ with $A \neq a_1$. EMD between the two groups represents the smallest amount of required change in the ranking scores of the bigger group to obtain the ranking scores of the smaller group. Then, EMD between G and $\bar{G}$ for

TABLE 3 Metrics categorized according to the fairness definition associated, with frequency of papers using the metric.

Fairness definition	Metric	Frequency
Demographic Parity (DP)	Earth Mover's Distance (EMD)	3
	Mann-Whitney U test (MWU)	3
	Normalized KL Divergence(NDKL)	2
	User Group Fairness (UGF)	1
	Salary Association Test (SAT)	1
	Set Difference Rate (SDR)	1
	List Difference Rate (LDR)	1
	Min/Max Skew (MS)	1
	Kendall's Tau (KT)	1
	Kruskal-Wallis H test (KWH)	1
	Total Difference between the Recall Curves (TDRC)	1
Equal opportunity (EO)	True Positive Rate Parity (TPRP)	4
	Adverse Imact Rate (AIR)	2
	Counterfactual Fraction (CF)	1
	True Negative Rate (TNR)	1
	Infeasible Index (II)	1
	Infeasible Count (IC)	1
	Exposure	1
Individual Fairness (IF)	Individual Fairness Certification (IFC)	1
	Effect Coefficient (EC)	1
Proportional Fairness (PF)	Violations	1
–	User Acceptance Test (UAT)	1
	Between Group Skill Difference (BGSD)	1
	Disparate Impact (DI)	1
	Statistical Parity (SP)	1

the smallest amount of change in the ranking score scenario is

$$EMD(G, \bar{G}) = \frac{\sum_{i,j} f_{i,j} \, |i - j|}{\sum_{i,j} f_{i,j}}$$

where $f_{i,j}$ is the change in i^{th} rank score of the bigger group to get the j^{th} rank score of the smaller group. To give an example, assume that the two groups are young G and old $\bar{G}$ candidates. Groups are represented by the score their members have at each rank. So, for $G = \{0, 0.3, 0.2\}$ and $\bar{G} = \{0.5, 0, 0\}$, the solution of smallest change is to rerank G to get the same score distribution as $\bar{G}$. Here, $EMD(G, \bar{G}) = \frac{0.3*1+0.2*2}{0.3+0.2} = 1.4$.

- *Normalized Discounted Kullback-Leibler Divergence (NDKL)* (Geyik et al., 2019): Given ranked list Q_u and two distributions $\mathcal{P}_{A_{Q_u}}$ and $\mathcal{P}_A$, which are distributions of attribute A in Q_u and the desired distribution of A, respectively, the NDKL is

$$NDKL(Q_u) = \frac{1}{Z} \sum_{j=1}^{|Q_u|} \frac{1}{log_2(j+1)} d_{KL}(\mathcal{P}_{A_{Q_u}} || \mathcal{P}_A),$$

$$Z = \sum_{j=1}^{|Q_u|} \frac{1}{log_2(j+1)}$$

where d_{KL} is KL-divergence. It should be noted that $NDKL = 0$ for all users would imply demographic parity is achieved.

- *Mann-Whitney U (MWU) test* (Corder and Foreman, 2014): MWU test is a non-parametric statistical test where the null hypothesis for job recommender is:
  - *For candidate recommender*: The ranks of candidates with a particular protected attribute is not significantly different from the ranks of candidates with another attribute.
  - *For job recommender*: The recommended list of job posts is not significantly different for candidates with different protected attributes.

The metrics associated with EO are dependent on the quality of recommendations and the distribution of sensitive attributes over recommended lists.

- *True Positive Rate Parity (TPRP)* (Delecraz et al., 2022): TPRP in candidate recommendation for a given binary protected attribute $A \in \{a_1, a_2\}$ and recommendation list Q_u for job u is defined as

$$TPRP(u) = \big|P(x \in Q_u | A_x = a_1, \rho(u,x) = 1) \\ - P(x \in Q_u | A_x = a_2, \rho(u,x) = 1)\big|$$

where $\rho(u, x) = 1$ implies that a candidate x sampled randomly from the candidate set is suitable for job u, and $TPRP = 0$ implies EO is achieved. TRPR is also called sourcing bias (Syed and Shivendu, 2022) and true positive rate gap (Jourdan et al., 2023) in the literature.

- *Adverse Impact Ratio (AIR)* (Burke et al., 2021): AIR for binary sensitive attribute $A \in \{a_1, a_2\}$ and recommended list Q_u for job u is,

$$AIR(u) = \frac{|P(x \in Q_u | A_x = a_1, \rho(u,x) = 1)|}{|P(x \in Q_u | A_x = a_2, \rho(u,x) = 1)|}$$

where, $\rho(u, x) = 1$ implies that a candidate x sampled randomly from the candidate set is suitable for job u, and $AIR = 1$ implies EO is achieved.

There also exist a few metrics which are not associated with a particular fairness definition but are nevertheless related with fairness, e.g., UAT measures the disadvantaged groups' perception of the RRS, and several metrics quantify the fairness of RRS datasets.

- *User Acceptance Test (UAT)* (Shishehchi and Banihashem, 2019): UAT is a questionnaire to assess the quality of job recommenders for people with disability, used by the authors in a user study. The questionnaire investigates four factors, i.e., usefulness, ease of use, ease of learning, and satisfaction with the RRS. The responses are registered in terms of values from 1 (disagree) to 5 (agree) for questions related to each factor.
- *Dataset Bias Metrics* (Delecraz et al., 2022): Statistical parity (SP) and Disparate impact (DI) metrics are used for measuring the bias in datasets with respect to binary attribute $A \in \{a_1, a_2\}$. Here, $\rho(u, x) = 1$ denotes that candidate u is qualified for a job post x randomly sampled from the set of job posts:

$$SP(u) = P(\rho(u,x) = 1 | A = a_1) - P(\rho(u,x) = 1 | A = a_2))$$

$$DI(u) = \frac{P(\rho(u,x) = 1 | A = a_1)}{P(\rho(u,x) = 1 | A = a_2)}$$

2.5. Investigating fairness of recommendations

A subset of reviewed articles analyzes fairness in RRSs. Tang et al. (2017) examines 17 million LinkedIn job listings spanning over 10 years and conducts a user study to analyze perceived stereotypes in these listings. The authors use the recruitment assistance services company Textio[9] and Unitive (no longer operational) to get a list of gendered words and then, based on the weighted frequency of those words in a job listing, measure their "maleness" and "femaleness." They find that job listings perceived as overall male and the usage of gendered words, in general, have decreased over the years. These results could suggest that our society is moving toward more gender-appropriate language. They also conducted a user study where two of the questions are "While reading the job description, to what extent did you feel that the advertisement would attract more male or more female applicants?" and "If you were fully qualified to apply for a job like this, how likely is it that you would apply for this particular position?" They compare user responses with the "genderedness" of job listings measured earlier using the Mann-Whitney U (MWU) test and found that there is a low correlation between the gendered wording of job listings and perceived gender bias (attractiveness to female applicants). Instead, the perceived bias depends on preconceived notions like technology jobs are male jobs or lower wage jobs are female jobs. Additionally, the willingness to apply had a low

9 https://textio.com/ (accessed May 2023).

correlation with perceived gender bias or the gendered wording of job listing.

Similarly, Chen et al. (2018) investigates gender bias on Indeed, Monster, and CareerBuilder resume searches. They use a regression model to measure IF and DP using MWU. The IF for the candidates occurring at ranks $30 - 50$ in the recommendation list shows that men occupy higher ranks compared to women, which can seem counter-intuitive. For DP, this advantage is significant for multiple job titles (e.g., Truck Driver and Software Engineer).

The work of Delecraz et al. (2022) analyzes bias of different attributes like age, gender, geography, and education in their private dataset with disparate impact (DI) and statistical parity (SP). They found that their dataset is fair along gender and age attributes but is unfair according to nationality. They use true positive rate parity (TPRP) to analyze EO over their private candidate recommender. They found that education, birthplace, and residence permit were impactful for the candidate selection of individuals, while age and gender were not. This is partially explained by the fact that education, birthplace, and residence permit, in many cases, can be seen as requirements rather than biases.

Elbassuoni et al. (2019, 2020) propose new heuristic- and decision-tree-based approaches, respectively, to find a partition of candidates based on their attributes for which unfairness in terms of DP is maximized. After partitioning, the authors use EMD to measure the unfairness in the ranked list. Then, following a similar methodological approach for measuring bias, Amer-Yahia et al. (2020) investigate the online recruitment platform TaskRabbit and the job recommender platform Google Job Search. The uniqueness of Elbassuoni et al. (2019, 2020) and Amer-Yahia et al. (2020) is that the authors consider partitioning based on combinations of attributes rather than single attributes.

Wilson et al. (2021) conduct a fairness audit of the Pymetrics candidate screening system.[10] Through this audit, the authors try to verify Pymetrics' claim to abide the $4/5^{\text{th}}$ rule from the US Union Guidelines on Employee Selection Procedures (Cascio and Aguinis, 2001). According to the $4/5^{\text{th}}$ rule, if the selection rate of a group is less than $4/5^{\text{th}}$ of the highest group selection rate, then that group is adversely impacted. This rule closely aligns with the DP definition of fairness.

Zhang (2021) provides a gender fairness audit of four Chinese job boards. On these job boards, there are some job posts with explicit mention of preferred gender. The audit using list/set difference rate (LDR/SDR) showed the existence of gender bias in terms of quality of recommendation, and also differences in the wording of job ads recommended to males versus those recommended to female job seekers.

Markert et al. (2022) is the only article to pursue individual fairness (IF) and adapt the classification IF certification (IFC) process for ranking. The IF definition for candidate recommenders in Section 2.3 requires similar candidates to have similar ranks. A regression model is trained to predict a candidate's rank in the recommended list given by the candidate recommender. The authors formulate a mixed integer linear programming problem using the regression model and the similarity constraint to get upper and lower bounds for the output of the regression model, i.e., the candidate's rank.

10 https://www.pymetrics.ai/ (accessed May 2023).

2.6. Pre-, in-, and post-processing approaches for fairness

The approaches to achieve fairness in RRSs can be categorized into pre-, in-, and post-processing techniques.

Pre-processing approaches are applied to the training data of the RS. The following approaches are used in the literature on RRSs:

- *Balancing the dataset*: Balancing the training data with respect to the sensitive attribute. For instance, Arafan et al. (2022) create gender-balanced synthetic data using CTGAN (conditional tabular generative adversarial network) for training (Xu et al., 2019), resulting in a significant decrease in NDKL.
- *Replacing the pronouns*: A simple approach for gender bias mitigation is to replace gendered pronouns with gender-neutral pronouns, as performed for instance in Rus et al. (2022) and Jourdan et al. (2023). This approach shows no change in terms of TPRP compared to not using pronoun substitution for Rus et al. (2022), while Jourdan et al. (2023) show a slight improvement in TPRP. The difference in results could be due to the different language models used.
- *Constrained resume sourcing*: Syed and Shivendu (2022) find conditions regarding the number of relevant candidates in each subgroup at the data sourcing of resumes for training to reduce the TRPR and theoretically achieve EO and DP.
- *Special group RSs*: The ontology-based/KB job recommenders use the sensitive information (e.g., disability, age, location, language) as input to create dedicated RSs for special groups (e.g., migrants or disabled people) (Shishehchi and Banihashem, 2019; Ntioudis et al., 2022). Shishehchi and Banihashem (2019) show strong acceptance of their RS by the special group (disabled people) using UAT, while Ntioudis et al. (2022) do not evaluate their system for the special group (migrant people).

In-processing approaches to mitigate bias change the RS itself to make the recommendations less biased. The in-processing approach is the least used method in the RRS literature. Also, in-processing approaches are only used with CBF.

- *Adversarial debiasing*: Rus et al. (2022) fine-tune a large language model to learn job-candidate similarity with the additional objective of removing gender information from the embedding of job posts, using an adversarial network that tries to predict the gender of the candidate. The adversarial debiasing shows significant improvement in TPRP compared to replacing pronouns.
- *Regularization-based debiasing*: Methods such as Jourdan et al. (2023) use a regularization term in the loss. network. In this case, the regularization term is the Sinkhorn Divergence (Chizat et al., 2020) over the distribution of sensitive attributes in the predictions. Similar to adversarial debiasing, this

approach shows significant improvement in TPRP compared to replacing pronouns.

Post-processing approaches are applied to the ranking received from the RRS to re-rank the items (jobs or candidates in our case). In research on fairness of RRSs, post-processing approaches are the most common, then pre-processing approaches, and in-processing are least explored in the literature (see Table 2), which is strikingly different from the fairness research in RSs overall (Deldjoo et al., 2023), where this order is in-, post-, and pre-processing, respectively. This comparison should be seen with caution though, as the number of papers surveyed is significantly less here compared to Deldjoo et al. (2023). The following are the post-processing approaches used in the RRSs we identified:

- *Introducing proportional fairness constraints*: Bei et al. (2020) try to achieve PF by removing the lowest ranking candidate inside the recommended list for which the attributes' upper bound condition (i.e., β) of PF gets violated, and adding the highest ranking candidate outside the recommended list for which the attributes' lower bound condition (i.e., α) of PF gets violated. This approach is able to approach PF with very few constraints violated.
- *Deterministic constrained sorting*: Geyik et al. (2019) introduce a deterministic sorting algorithm with constraints similar to the PF constraints. The difference with the work of Bei et al. (2020) is that here each candidate has only one attribute rather than a set of attributes and the size of the recommended list is not fixed here. Arafan et al. (2022) re-rank the recommended list to achieve the same number of candidates for each attribute using the re-ranking algorithm of Geyik et al. (2019). Both papers show improvement in NDKL. Arafan et al. (2022) additionally show that the NDKL scores of the deterministic constrained sorting approach can be further improved by using an artificially balanced dataset.
- *Spatial partitioning*: Burke et al. (2021) introduce a re-ranking algorithm based on spatial partitioning of 3-dimensional space created by three attributes of the candidate, i.e., performance, retention after hiring, and whether the candidate belongs to a protected group or not. Improvements in terms of AIR score are mentioned by Burke et al. (2021) though scores are not reported.
- *Targeting user group fairness*: Li et al. (2023) re-rank the recommended list of jobs for all candidates to maximize the sum of each user's personalized utility score (Zhang et al., 2016) over all candidate-job pairs while minimizing the UGF value.
- *Intervention-based skill improvement*: The method proposed by Scher et al. (2023) selects candidates to upgrade their skills so that it improves their probability of selection by the candidate recommenders. The authors divide the candidates into high-prospect group and low-prospect group using their skills and sensitive information as decision criteria. Then, the high-prospect groups' skill is upgraded and the skill upgrade of the low-prospect group is delayed for some time. The method helps the high-prospect group and punishes the low-prospect group, resulting in less reduction of inequality (i.e., less improvement in EO and DP) in the long run compared to random selection of candidates. And the selection of only low-prospect group has no impact on EO in the long run.
- *Attribute intervention*: Tran et al. (2021) and Tran et al. (2022) identify the skill set of the candidate that should be upgraded and the level to which it should be upgraded based on candidate attributes (including protected attributes) to achieve improvement in selection probability by a causal tree-based candidate recommender, more precisely, by a maximal causal tree (Tran et al., 2021) or personalized causal tree (Tran et al., 2022).

The reviewed debiasing approaches work toward fairer RRSs from a technical perspective, but the recruitment domain requires to take a legal perspective too since obligations and requirements from employment law apply. In the next section, we will therefore give an overview of the most critical legal pitfalls for RRSs. With this, we aim at giving the researchers and developers of RRSs guidelines on possible legal issues of their systems.

3. Legal validity

When RSs are used in the hiring process to (help) make decisions, the legal requirements from employment law are applicable to them in a similar manner as to human decision-makers (Barocas et al., 2019). In the case of hiring decisions, with or without algorithmic support, *non-discrimination law* (European Council, Directive 2000/43/EC, 2000; European Council, Directive 2000/78/EC, 2000; European Parliament and the Council, Directive 2006/54/EC, 2006) and *data protection law* [European Parliament and the Council, Regulation (EU) 2016/679, 2016] in particular must be observed (Hacker, 2018). RS users, in addition, should also take note of newly adopted laws regulating the online realm and AI technology in the EU.[11] Most notably in this context are the Digital Markets Act [European Parliament and the Council, Regulation (EU) 2022/1925, 2022] and the Digital Services Act [European Parliament and the Council, Regulation (EU) 2022/2065, 2022] which are already in force, as well as the Proposal for an Artificial Intelligence Act (European Commission, Proposal Artificial Intelligence Act, 2021) which could become law within 2024. While ordinarily the term "fairness" is used to describe fair and equal divisions of resources, in law the concept dealing with this in the world of employment is "non-discrimination." The fundamental right of non-discrimination is highlighted in all of the recent EU regulations. In the specific case of RRSs, by virtue of being applied in recruitment, the algorithmic decision has to comply with the directives of non-discrimination law.

Compared to these legal sources and terminologies, researchers and practitioners in computer science (CS) and artificial intelligence (AI) commonly use the term "fairness" and quantify it according to some computational metric, as we have seen in Sections 2.3 and 2.4. Therefore, their assumption is that a system can be more or less fair. In stark contrast, the legality of a

11 https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence (accessed May 2023).

decision cannot be measured, it is either a case of discrimination or it is not. The benchmark is non-discrimination law. A system cannot in principle base its hiring decision on the sex, race or ethnicity, religion or belief, disability, age, or sexual orientation of the candidates, as EU law prohibits unjustified differential treatment on the basis of these protected characteristics (direct discrimination). A system can also not use an apparently neutral criterion which will have the effect of disadvantaging a considerably higher percentage of persons sharing the protected characteristic (indirect discrimination). This would happen if a system is applied in the same way to everybody, but disadvantages a group of people who share a protected characteristic.

The legal literature, despite the terminological incongruence, has picked up the most used fairness definitions in CS and AI research and often categorized algorithms that adopt them as either blind (or unaware) algorithms or protected-characteristics-aware algorithms (Žliobaitė and Custers, 2016; Bent, 2019; Wachter et al., 2020; Kim, 2022). The former describes algorithmic designs that achieve their results without using any of the protected characteristics as grounds for decision. The algorithm is, therefore, not given any information about any of the protected attributes. The latter describes algorithms which use some or all protected characteristics in their data to make decisions. Blind algorithms not only tend to underperform (Žliobaitė and Custers, 2016; Bent, 2019; Xiang and Raji, 2019; Wachter et al., 2020; Kim, 2022), but also often learn spurious correlations in the data that can serve as proxies for the protected characteristics in their datasets, which can equally lead to discrimination cases (Žliobaitė and Custers, 2016; Chander, 2017; Kim, 2017; Bent, 2019; Wachter et al., 2020; Adams-Prassl, 2022; Hildebrandt, 2022).[12] This can occur, for instance, when a correlation between ethnicity and postal code is drawn or when the recommendation algorithm unintentionally incorporates implicit gender information from interaction data because of different preferences of male and female users (Ganhör et al., 2022). The fairness definitions presented above all use protected characteristics to achieve fairness (Barocas et al., 2019; Bent, 2019).

3.1. Are debiased candidate recommenders affirmative action measures?

Debiased candidate recommenders essentially re-rank candidates according to the fairness definition used in order to achieve a parity for groups with and without certain protected characteristics. Job recommenders inherently also perform a re-ranking, however of job advertisement and employment opportunities. The discrimination risks thereby are similar, albeit center more on the delivery system (e.g., targeted ads). A closer investigation though transcends the scope of this survey (Greif and Grosz, 2023). Affirmative action schemes, known as "positive action" in EU law, are measures by which specific advantages are given to the underrepresented groups in order to compensate for existing disadvantages in working life to ensure full equality. Quotas are the most used example and directly relate to what RRSs are doing by re-ranking candidates. "Fair" algorithmic affirmative action, as strived for by the fairness definitions, does however not translate to a lawful understanding of the concept (Xiang and Raji, 2019). The Court of Justice of the EU (CJEU) has previously stated that schemes which give "absolute and unconditional priority" exceed the limitations of the positive action exception (CJEU C-450/93, 1995). In subsequent case-law, the CJEU however narrowed the scope explaining that flexible quotas allowing for individual consideration would be in line with the positive action exception (CJEU C-409/95, 1997). As long as a "saving clause" is provided for, allowing for an objective assessment of all criteria, which can "override the priority accorded [...] where one or more of those criteria tilts the balance in favor of [another] candidate," a quota scheme would be line with EU law (CJEU C-409/95, 1997). Thus, every case involving the hiring according to a quota scheme must be open for individual consideration. Whether this could be computable is doubtful (Hacker, 2018; Adams-Prassl, 2022). Any candidate recommender re-ranking candidates to achieve a fixed fairness definition without taking each candidates' individual circumstances into account will most likely run afoul of the legal requirements set out. For instance, imagine an employer deciding between two candidates, A and B, to fill one position. A and B are materially equally qualified, meaning that there resumes, though not necessarily the same, are of equivalent value for this position. Candidate A should be hired according to the affirmative action scheme in place. The employer however should (according to the "savings clause") still hire candidate B, if there are individual circumstances, reasons specific to that candidate (e.g., sole provider or long term unemployed), which tilts the balance in candidate B's favor. This needs to be decided on an individual case level and is therefore hard to automatize in a RRSs. This, however, must not mean the end of all algorithmically assisted hiring.

3.2. Pre-, in-, or post-processing: walking the legal line

As Hacker (2018) notes, the case-law dealing essentially with "corrective powers" at or after the selection process faces greater scrutiny from the CJEU than measures applied before. On the other hand, the CJEU seems to be more lenient when positive action measures are applied *before* first selection (CJEU C-158/97, 2000). Indeed, "positive measures" (even strict quotas) before the actual selection stage of the decision procedure are more likely to be accepted by the CJEU (Hacker, 2018). This could make implementing algorithmic fairness during the *training* stage of a model and before actually ranking candidates wrapped as a quota scheme widely applicable (Hacker, 2018; Adams-Prassl, 2022). Approaches including balancing the training dataset or re-ranking of (fictitious) candidates in the training phase of a candidate recommender should in principle be valid options for routing out biases before the model is put on the market. A glance over to the US shows a similar approach: US case-law suggests that rearrangement in terms of affirmative action applied *after* the selection results have been allocated to the respective

12 Proxy discrimination most often comes in the form of indirect discrimination.

candidates[13] could lead to discrimination of the now down-ranked selected candidates (US Supreme Court, Ricci v DeStefano, 2009).

For both jurisdictions (EU and US), the crux is timing: taking into account biases (and potentially discrimination) that is found in society and in datasets is possible (also via algorithmic help), as long as it is done *before* real-world application. Otherwise, a candidate recommender risks producing further discrimination by ranking candidates first and foremost according to a protected characteristic (e.g., sex) and not according to actual suitability for the job in question. Kim (2022) notes additionally that the US Supreme Court acknowledges that an employer may need to take protected characteristics into account to create fairer hiring processes. Along with Bent (2019), Kim (2017) argues that this should leave room for algorithmic affirmative action before the selection process. Therefore, as it stands, in terms of (both EU and US) non-discrimination law, pre-processing approaches should be favored, whereas post- and probably in-processing approaches would most likely run afoul of current requirements.

4. Conclusions and open challenges

In this survey, we analyzed the current fairness research in RRSs from multiple perspectives. First, the algorithms used in RRSs were classified into four categories. Subsequently, we consolidated different fairness definitions in the existing literature to understand the objectives of fairness research in RRSs. We provided the fairness definitions used in classification and connected them with their adapted forms for RRSs. Further, we detailed some of the fairness metrics found in the surveyed literature. We also discussed the work done to analyze the presence or absence of fairness in RRSs. Subsequently, the most common pre-, in-, and post-processing approaches to gain fairness in RRSs were described. Finally, we bridge the gap to legal scholarship by discussing fairness definitions and their relation to legal requirements in order to provide an overview of some of the possible legal issues resulting from unfairness in RRSs. We thereby identified the lack of interdisciplinary vocabulary and understanding as a substantial challenge.

RRSs is a quickly evolving field but is also facing several challenges and open questions which are waiting to be solved. First, the *lack of public datasets for fairness research* in the recruitment domain is evident, as highlighted in Table 2. The usage of private data and proprietary RRSs limits the understanding and reproducibility of fairness research.

Experiments are most often *limited to gender* as sensitive attribute. Other attributes such as ethnicity, age, or disability are commonly only targeted when studied in conjunction with gender and are often modeled artificially. This limits our understanding of the role of non-gender attributes in the recruitment domain. At the same time, it is exciting to see the recent works considering discrimination against groups defined by not only a single attribute but a combination of attributes.

Fairness definitions are also an increasing challenge for the community. They are most commonly *adapted from classification fairness definitions* for RSs. The adapted definitions for job recommenders does not consider the ranks of recommended items. This, however, is important for both candidate and job recommenders. Individual fairness (Markert et al., 2022) is an exciting new direction away from the standard of group-associated fairness definitions such as demographic parity and equal opportunity.

The fairness metrics are highly diverse across the literature surveyed. We would like to highlight here that even when the fairness definitions targeted are the same, the metrics used for fairness measurement are rarely identical. This points to the problem of a *lack of standardized fairness evaluation metrics* in the recruitment domain.

The recruitment domain is a content-rich domain (i.e., resumes and job posts both convey lots of descriptive textual semantics), which explains the prevalence of CBF and KB recommendation algorithms. The *scarcity of hybrid and CF algorithms* in RRSs research on fairness shows a disconnect with common RSs research (de Ruijt and Bhulai, 2021). In addition, while currently *post-processing approaches are most often adopted* for debiasing, as discussed in Sections 2.6, 3.2, they are also the most concerning ones from a legal perspective. We, therefore, suggest to devote more research to pre-processing and in-processing strategies. In addition, we strongly advocate for more interdisciplinary research, involving experts from both RSs and legal scholarship, to formulate strategies and constraints for legally suitable in- and post-processing approaches and to drive RS research in the recruitment domain.

To wrap up, the surveyed research works are (1) diverse in terms of job/candidate recommenders and the adopted algorithms, (2) explore new fairness definitions such as IF, and (3) experiment with various metrics that are attempts to better represent the same fairness concept. The current trajectory of fairness research in RRSs is highly promising, but several avenues for further improvements through the valuable input from diverse research communities is required.

Author contributions

Conceptualization and funding acquisition: MS, EG, and NR. Literature search and writing—original draft: DK and TG. Writing—review and editing: MS, NG, and EG. All authors approved the submitted version.

Funding

This research was funded by the Austrian Science Fund (FWF): DFH-23 and P33526 and by the State of Upper Austria and the

13 In the case of RRSs, this means that candidates have received their score from the system, based on which the ranked list is created.

Federal Ministry of Education, Science, and Research, through grants LIT-2020-9-SEE-113 and LIT-2021-YOU-215.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abel, F., Benczúr, A., Kohlsdorf, D., Larson, M., and Pálovics, R. (2016). "Recsys challenge 2016: Job recommendations," in *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16* (New York, NY, USA: Association for Computing Machinery), 425–426. doi: 10.1145/2959100.2959207

Abel, F., Deldjoo, Y., Elahi, M., and Kohlsdorf, D. (2017). "Recsys challenge 2017: Offline and online evaluation," in *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys '17* (New York, NY, USA: Association for Computing Machinery), 372–373. doi: 10.1145/3109859.3109954

Adams-Prassl, J. (2022). Regulating algorithms at work: Lessons for a "European approach to artificial intelligence." *Eur. Labour Law J.* 13, 30–50. doi: 10.1177/20319525211062558

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., and Rieke, A. (2019). Discrimination through optimization: How Facebook's ad delivery can lead to biased outcomes. *Proc. ACM Hum.-Comput. Interact.* 3, 1–30. doi: 10.1145/3359301

Amer-Yahia, S., Elbassuoni, S., Ghizzawi, A., Borromeo, R. M., Hoareau, E., and Mulhem, P. (2020). "Fairness in online jobs: A case study on TaskRabbit and Google," in *International Conference on Extending Database Technologies (EDBT)* (Copenhagen, Denmark).

Arafan, A. M., Graus, D., Santos, F. P., and Beauxis-Aussalet, E. (2022). "End-to-end bias mitigation in candidate recommender systems with fairness gates," in *2nd Workshop on Recommender Systems for Human Resources, RecSys-in-HR 2022. CEUR-WS.*

Barocas, S., Hardt, M., and Narayanan, A. (2019). *Fairness and Machine Learning: Limitations and Opportunities*. Cambridge, MA, MIT press.

Becker, B., and Kohavi, R. (1996). Adult. *UCI Machine Learning Repository*. Available online at: https://archive-beta.ics.uci.edu/dataset/2/adult (accessed September 27, 2023).

Bei, X., Liu, S., Poon, C. K., and Wang, H. (2020). "Candidate selections with proportional fairness constraints," in *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '20* (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), 150–158.

Bent, J. R. (2019). Is algorithmic affirmative action legal. *Geo. LJ* 108, 803.

Bian, S., Chen, X., Zhao, W. X., Zhou, K., Hou, Y., Song, Y., et al. (2020). "Learning to match jobs with resumes from sparse interaction data using multi-view co-teaching network," in *Proceedings of the 29th ACM International Conference on Information Knowledge Management, CIKM '20* (New York, NY, USA: Association for Computing Machinery), 65–74. doi: 10.1145/3340531.3411929

Burke, I., Burke, R., and Kuljanin, G. (2021). "Fair candidate ranking with spatial partitioning: Lessons from the SIOP ML competition," in *Proceedings of the First Workshop on Recommender Systems for Human Resources (RecSys in HR 2021) co-located with the 15th ACM Conference on Recommender Systems (RecSys 2021)* (CEUR-WS).

Cascio, W. F., and Aguinis, H. (2001). The federal uniform guidelines on employee selection procedures (1978), an update on selected issues. *Rev. Public Person. Admin.* 21, 200–218. doi: 10.1177/0734371X0102100303

Chander, A. (2017). The racist algorithm? *Michigan Law Rev.* 115, 1023. doi: 10.36644/mlr.115.6.racist

Chen, L., Ma, R., Hannák, A., and Wilson, C. (2018). "Investigating the impact of gender on rank in resume search engines," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18* (New York, NY, USA: Association for Computing Machinery), 1–14. doi: 10.1145/3173574.3174225

Chizat, L., Roussillon, P., Léger, F., Vialard, F.-X., and Peyré, G. (2020). "Faster Wasserstein distance estimation with the Sinkhorn divergence," in *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS'20* (Red Hook, NY, USA: Curran Associates Inc.).

CJEU C-158/97 (2000). Georg Badeck and Others v Landesanwalt beim Staatsgerichtshof des Landes Hessen.

CJEU C-409/95 (1997). Hellmut Marschall v Land Nordrhein-Westfalen.

CJEU C-450/93 (1995). Eckhard Kalanke v Freie Hansestadt Bremen.

Corder, G. W., and Foreman, D. I. (2014). *Nonparametric Statistics: A Step-by-Step Approach*. New York, NY: John Wiley Sons.

Dastin, J. (2018). "Amazon scraps secret AI recruiting tool that showed bias against women," in *Ethics of Data and Analytics* (Auerbach Publications), 296–299. doi: 10.1201/9781003278290-44

de Ruijt, C., and Bhulai, S. (2021). Job recommender systems: A review. *arXiv preprint arXiv:2111.13576.*

De-Arteaga, M., Romanov, A., Wallach, H., Chayes, J., Borgs, C., Chouldechova, A., et al. (2019). "Bias in bios: A case study of semantic representation bias in a high-stakes setting," in *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19* (New York, NY, USA: Association for Computing Machinery), 120–128. doi: 10.1145/3287560.3287572

Deldjoo, Y., Jannach, D., Bellogin, A., Difonzo, A., and Zanzonelli, D. (2023). Fairness in recommender systems: Research landscape and future directions. *User Model User-Adap. Inter.* 30, 1–50. doi: 10.1007/s11257-023-09364-z

Delecraz, S., Eltarr, L., Becuwe, M., Bouxin, H., Boutin, N., and Oullier, O. (2022). "Making recruitment more inclusive: Unfairness monitoring with a job matching machine-learning algorithm," in *Proceedings of the 2nd International Workshop on Equitable Data and Technology, FairWare '22* (New York, NY, USA: Association for Computing Machinery), 34–41. doi: 10.1145/3524491.3527309

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (Minneapolis, MA: Association for Computational Linguistics), 4171–4186.

Dhameliya, J., and Desai, N. (2019). "Job recommender systems: A survey," in *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)* 1–5. doi: 10.1109/i-PACT44901.2019.8960231

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. (2012). "Fairness through awareness," in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, ITCS '12* (New York, NY, USA: Association for Computing Machinery), 214–226. doi: 10.1145/2090236.2090255

Elbassuoni, S., Amer-Yahia, S., and Ghizzawi, A. (2020). Fairness of scoring in online job marketplaces. *ACM/IMS Trans. Data Sci.* 1, 1–30. doi: 10.1145/3402883

Elbassuoni, S., Amer-Yahia, S., Ghizzawi, A., and Atie, C. E. (2019). "Exploring fairness of ranking in online job marketplaces," in *22nd International Conference on Extending Database Technology (EDBT)* (Lisbone, Portugal).

European Commission, Proposal Artificial Intelligence Act (2021). Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts: COM 2021 206 final.

European Council, Directive 2000/43/EC (2000). Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin: Race Directive.

European Council, Directive 2000/78/EC (2000). Directive 2000/78/EC establishing a general framework for equal treatment in employment and occupation: Framework Directive.

European Parliament and the Council, Directive 2006/54/EC (2006). Directive 2006/54/EC on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast): Gender Recast Directive.

European Parliament and the Council, Regulation (EU) 2016/679 (2016). The Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive (EC) 95/46 (General Data Protection Regulation): GDPR.

European Parliament and the Council, Regulation (EU) 2022/1925 (2022). Regulation (EU) 2022/1925 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act): DMA.

European Parliament and the Council, Regulation (EU) 2022/2065 (2022). Regulation (EU) 2022/2065 on a Single Market For Digital Services and amending Directive (EC) 2000/31 (Digital Services Act): DSA.

Fawaz, N. (2019). Recommending dream jobs in a biased real world. *arXiv preprint arXiv:1905.06134*.

Freire, M. N., and de Castro, L. N. (2021). E-recruitment recommender systems: A systematic review. *Knowl. Inf. Syst*. 63, 1–20. doi: 10.1007/s10115-020-01522-8

Ganhör, C., Penz, D., Rekabsaz, N., Lesota, O., and Schedl, M. (2022). "Unlearning protected user attributes in recommendations with adversarial training," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22* (New York, NY, USA: Association for Computing Machinery), 2142–2147.

Geyik, S. C., Ambler, S., and Kenthapadi, K. (2019). "Fairness-aware ranking in search recommendation systems with application to LinkedIn talent search," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining, KDD '19* (New York, NY, USA: Association for Computing Machinery), 2221–2231. doi: 10.1145/3292500.3330691

Greif, E., and Grosz, T. (2023). To see, or Not to see: Online job advertisement and EU non-discrimination law. *Eur. Labour Law J*. 14, 20319525231172089. doi: 10.1177/20319525231172089

Gugnani, A., and Misra, H. (2020). "Implicit skills extraction using document embedding and its use in job recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence* 13286–13293. doi: 10.1609/aaai.v34i08.7038

Gutiérrez, F., Charleer, S., De Croon, R., Htun, N. N., Goetschalckx, G., and Verbert, K. (2019). "Explaining and exploring job recommendations: A user-driven approach for interacting with knowledge-based job recommender systems," in *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys '19* (New York, NY, USA: Association for Computing Machinery), 60–68. doi: 10.1145/3298689.3347001

Hacker, P. (2018). Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. *Common. Market Law Rev*. 55, 1143–1186. doi: 10.54648/COLA2018095

Hardt, M., Price, E., Price, E., and Srebro, N. (2016). "Equality of opportunity in supervised learning," in *Advances in Neural Information Processing Systems*, eds. D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Red Hook, NY: Curran Associates, Inc.).

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., and Chua, T.-S. (2017). "Neural collaborative filtering," in *Proceedings of the 26th International Conference on World Wide Web, WWW '17* (Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee), 173–182. doi: 10.1145/3038912.3052569

Hildebrandt, M. (2022). The Issue of Proxies and Choice Architectures. Why EU law matters for recommender systems. *Front. Artif. Intell*. 73, 789076. doi: 10.3389/frai.2022.789076

Jourdan, F., Kaninku, T. T., Asher, N., Loubes, J.-M., and Risser, L. (2023). How optimal transport can tackle gender biases in multi-class neural network classifiers for job recommendations. *Algorithms* 16, 174. doi: 10.3390/a16030174

Kim, P. T. (2017). Data-driven discrimination at work. *Wm. Mary L. Rev*. 58, 857.

Kim, P. T. (2022). Race-aware algorithms: Fairness, nondiscrimination and affirmative action. *Cal. L. Rev*. 110, 1539. doi: 10.15779/Z387P8TF1W

Koenig, N., and Thompson, I. (2021). "The 2020–2021 SIOP machine learning competition," in *36th annual Society for Industrial and Organizational Psychology Conference* (New Orleans, LA: SIOP).

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer* 42, 30–37. doi: 10.1109/MC.2009.263

Lacic, E., Reiter-Haas, M., Kowald, D., Reddy Dareddy, M., Cho, J., and Lex, E. (2020). Using autoencoders for session-based job recommendations. *User Model. User-Adapt. Inter*. 30, 617–658. doi: 10.1007/s11257-020-09269-1

Li, Y., Yamashita, M., Chen, H., Lee, D., and Zhang, Y. (2023). "Fairness in job recommendation under quantity constraints," in *AAAI-23 Workshop on AI for Web Advertising*.

Liu, Q., Zeng, Y., Mokhosi, R., and Zhang, H. (2018). "STAMP: Short-term attention/memory priority model for session-based recommendation," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery Data Mining, KDD '18* (New York, NY, USA: Association for Computing Machinery), 1831–1839. doi: 10.1145/3219819.3219950

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *CoRR, abs/1907.11692*.

Luo, Y., Zhang, H., Wen, Y., and Zhang, X. (2019). "ResumeGAN: An optimized deep representation learning framework for talent-job fit via adversarial learning," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19* (New York, NY, USA: Association for Computing Machinery), 1101–1110. doi: 10.1145/3357384.3357899

Markert, K., Ahouzi, A., and Debus, P. (2022). Fairness in regression-analysing a job candidates ranking system. *INFORMATIK* 2022, 1275–1285. doi: 10.18420/inf2022_109

Ntioudis, D., Masa, P., Karakostas, A., Meditskos, G., Vrochidis, S., and Kompatsiaris, I. (2022). Ontology-based personalized job recommendation framework for migrants and refugees. *Big Data Cogn. Comput*. 6, 120. doi: 10.3390/bdcc6040120

Pele, O., and Werman, M. (2009). "Fast and robust earth mover's distances," in *2009 IEEE 12th International Conference on Computer Vision* 460–467. doi: 10.1109/ICCV.2009.5459199

Rafter, R., Bradley, K., and Smyth, B. (2000). "Personalised retrieval for online recruitment services," in *The BCS/IRSG 22nd Annual Colloquium on Information Retrieval (IRSG 2000)* (Cambridge, UK).

Raghavan, M., Barocas, S., Kleinberg, J., and Levy, K. (2020). "Mitigating bias in algorithmic hiring: Evaluating claims and practices," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20* (New York, NY, USA: Association for Computing Machinery), 469–481. doi: 10.1145/3351095.3372828

Ruoss, A., Balunović, M., Fischer, M., and Vechev, M. (2020). "Learning certified individually fair representations," in *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS'20* (Red Hook, NY, USA: Curran Associates Inc.).

Rus, C., Luppes, J., Oosterhuis, H., and Schoenmacker, G. H. (2022). "Closing the gender wage gap: Adversarial fairness in job recommendation," in *2nd Workshop on Recommender Systems for Human Resources, RecSys-in-HR 2022. CEUR-WS*.

Salakhutdinov, R., and Mnih, A. (2007). "Probabilistic matrix factorization," in *Proceedings of the 20th International Conference on Neural Information Processing Systems, NIPS'07* (Red Hook, NY, USA: Curran Associates Inc.), 1257–1264.

Scher, S., Kopeinik, S., Trügler, A., and Kowald, D. (2023). Modelling the long-term fairness dynamics of data-driven targeted help on job seekers. *Scient. Rep*. 13, 1727. doi: 10.1038/s41598-023-28874-9

Shishehchi, S., and Banihashem, S. Y. (2019). JRDP: A job recommender system based on ontology for disabled people. *Int. J. Technol. Hum. Interact*. 15, 85–99. doi: 10.4018/IJTHI.2019010106

Smith, S. M., Edwards, R., and Duong, H. C. (2021). Unemployment rises in 2020, as the country battles the COVID-19 pandemic. *Monthly Lab. Rev*. 144, 1. doi: 10.21916/mlr.2021.12

Syed, R. A., and Shivendu, S. (2022). *Sourcing bias on job matching platforms: Regulation and optimal platform strategy*. Available online at: https://ssrn.com/abstract=4188162 (accessed June 15, 2022). doi: 10.2139/ssrn.4188162

Tang, S., Zhang, X., Cryan, J., Metzger, M. J., Zheng, H., and Zhao, B. Y. (2017). Gender bias in the job market: A longitudinal analysis. *Proc. ACM Hum.-Comput. Interact*. 1, 1–19. doi: 10.1145/3134734

Thali, R., Mayekar, S., More, S., Barhate, S., and Selvan, S. (2023). "Survey on job recommendation systems using machine learning," in *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)* 453–457. doi: 10.1109/ICIDCA56705.2023.10100122

Tran, H. X., Le, T. D., Li, J., Liu, L., Liu, J., Zhao, Y., et al. (2021). "Recommending the most effective intervention to improve employment for job seekers with disability," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery Data Mining, KDD '21* (New York, NY, USA: Association for Computing Machinery), 3616–3626. doi: 10.1145/3447548.3467095

Tran, H. X., Le, T. D., Li, J., Liu, L., Liu, J., Zhao, Y., et al. (2022). "Recommending personalized interventions to increase employability of disabled jobseekers," in *Advances in Knowledge Discovery and Data Mining: 26th Pacific-Asia Conference, PAKDD 2022, Chengdu, China, May 16–19, 2022, Proceedings, Part III* (Berlin, Heidelberg: Springer-Verlag), 92–104. doi: 10.1007/978-3-031-05981-0_8

US Supreme Court. (2009). *Ricci v DeStefano*. 557 *U.S*. 557 2009.

Vega, J. (1990). "Semantic matching between job offers and job search requests," in *COLING 1990 Volume 1: Papers presented to the 13th International Conference on Computational Linguistics. doi: 10.3115/992507.992520*

Wachter, S., Mittelstadt, B., and Russell, C. (2020). Bias preservation in machine learning: the legality of fairness metrics under EU non-discrimination law. *West Virginia Law Rev*. 123, 735. doi: 10.2139/ssrn.3792772

Wang, Y., Ma, W., Zhang, M., Liu, Y., and Ma, S. (2023). A survey on the fairness of recommender systems. *ACM Trans. Inf. Syst*. 41, 1–43. doi: 10.1145/3594633

Wilson, C., Ghosh, A., Jiang, S., Mislove, A., Baker, L., Szary, J., et al. (2021). "Building and auditing fair algorithms: A case study in candidate screening," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21* (New York, NY, USA: Association for Computing Machinery), 666–677. doi: 10.1145/3442188.3445928

Wilson, R., Beaven, R., May-Gillings, M., Hay, G., and Stevens, J. (2014). *Working Futures 2012–2022*. Evidence Report 83.

Xiang, A., and Raji, I. D. (2019). On the legal compatibility of fairness definitions. *arXiv preprint arXiv:1912.00761*. doi: 10.48550/arXiv.1912.00761

Xu, L., Skoularidou, M., Cuesta-Infante, A., and Veeramachaneni, K. (2019). "Modeling tabular data using conditional GAN," in *Advances in Neural Information Processing Systems*, eds. H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Red Hook, NY, USA: Curran Associates, Inc.).

Zehlike, M., Yang, K., and Stoyanovich, J. (2022). Fairness in ranking, Part II: Learning-to-rank and recommender systems. *ACM Comput. Surv*. 55, 1–41. doi: 10.1145/3533380

Zhang, S. (2021). *Measuring algorithmic bias in job recommender systems: An audit study approach*. Technical report. doi: 10.1257/rct.6101-1.1

Zhang, Y., Zhao, Q., Zhang, Y., Friedman, D., Zhang, M., Liu, Y., et al. (2016). "Economic recommendation with surplus maximization," in *Proceedings of the 25th International Conference on World Wide Web, WWW '16* (Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee), 73–83. doi: 10.1145/2872427.2882973

Žliobaitė, I., and Custers, B. (2016). Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models. *Artif. Intell. Law* 24, 183–201. doi: 10.1007/s10506-016-9182-5

frontiers | Frontiers in Big Data

TYPE Mini Review
PUBLISHED 12 October 2023
DOI 10.3389/fdata.2023.1249997

Check for updates

OPEN ACCESS

EDITED BY
Yassine Himeur,
University of Dubai, United Arab Emirates

REVIEWED BY
Chaochao Chen,
Zhejiang University, China

*CORRESPONDENCE
Peter Müllner
✉ pmuellner@know-center.at;
✉ pmuellner@student.tugraz.at
Elisabeth Lex
✉ elisabeth.lex@tugraz.at
Dominik Kowald
✉ dkowald@know-center.at

RECEIVED 29 June 2023
ACCEPTED 25 September 2023
PUBLISHED 12 October 2023

CITATION
Müllner P, Lex E, Schedl M and Kowald D (2023) Differential privacy in collaborative filtering recommender systems: a review.
Front. Big Data 6:1249997.
doi: 10.3389/fdata.2023.1249997

COPYRIGHT
© 2023 Müllner, Lex, Schedl and Kowald. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Differential privacy in collaborative filtering recommender systems: a review

Peter Müllner[1,2]*, Elisabeth Lex[2]*, Markus Schedl[3,4] and Dominik Kowald[1,2]*

[1]Know-Center Gmbh, Graz, Austria, [2]Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, Austria, [3]Institute of Computational Perception, Johannes Kepler University Linz, Linz, Austria, [4]Linz Institute of Technology, Linz, Austria

State-of-the-art recommender systems produce high-quality recommendations to support users in finding relevant content. However, through the utilization of users' data for generating recommendations, recommender systems threaten users' privacy. To alleviate this threat, often, differential privacy is used to protect users' data via adding random noise. This, however, leads to a substantial drop in recommendation quality. Therefore, several approaches aim to improve this trade-off between accuracy and user privacy. In this work, we first overview threats to user privacy in recommender systems, followed by a brief introduction to the differential privacy framework that can protect users' privacy. Subsequently, we review recommendation approaches that apply differential privacy, and we highlight research that improves the trade-off between recommendation quality and user privacy. Finally, we discuss open issues, e.g., considering the relation between privacy and fairness, and the users' different needs for privacy. With this review, we hope to provide other researchers an overview of the ways in which differential privacy has been applied to state-of-the-art collaborative filtering recommender systems.

KEYWORDS

differential privacy, collaborative filtering, recommender systems, accuracy-privacy trade-off, review

1. Introduction

Several previous research works have revealed multiple privacy threats for users in recommender systems. For example, the disclosure of users' private data to untrusted third parties (Calandrino et al., 2011), or the inference of users' sensitive attributes, such as gender or age (Zhang et al., 2023). Similarly, also the users themselves care more about their privacy in recommender systems (Herbert et al., 2021). For these reasons, privacy-enhancing techniques have been applied, most prominently *differential privacy (DP)* (Dwork, 2008). DP injects random noise into the recommender system and formally guarantees a certain degree of privacy. However, through this random noise, the quality of the recommendations suffers (Berkovsky et al., 2012). Many works aim to address this trade-off between recommendation quality and user privacy via carefully applying DP in specific ways. Friedman et al. (2016) show that in case of matrix factorization, DP can be applied to three different parts of the recommender system: (i) to the input of the recommender system, (ii) within the training process of the model, and (iii) to the model after training. However, a concise overview of works with respect to these three categories does not exist yet.

Therefore, in the paper at hand, we address this gap and identify 26 papers from relevant venues that deal with DP in collaborative filtering recommender systems. We briefly review these 26 papers and make two key observations about the state-of-the-art. Firstly, the vast majority of works use datasets from the same non-sensitive domain, i.e., movies. Secondly, research on applying DP after model training is scarce. Finally, we discuss our findings and present two open questions that may be relevant for future research: *How does applying DP impact fairness?* and *How to quantify the user's perceived privacy?*

Our work is structured as follows: In Section 2, we present threats to the privacy of users in recommender systems and additionally, introduce the DP framework. In Section 3, we precisely outline our methodology for obtaining the set of 26 relevant papers. In Section 4, we review these papers and group them into three groups according to the way in which they apply DP. In Section 5, we discuss our findings and propose open issues that we identified.

2. Background

In recent years, users of recommender systems have shown increasing concerns with respect to keeping their data private (Herbert et al., 2021). In fact, several research works (Bilge et al., 2013; Jeckmans et al., 2013; Friedman et al., 2015; Beigi and Liu, 2020; Majeed and Lee, 2020; Himeur et al., 2022) have revealed multiple privacy threats, for example, the inadvertent disclosure of users' interaction data, or the inference of users' sensitive attributes (e.g., gender, age).

Typically, a recommender system utilizes historic interaction data to generate recommendations. Ramakrishnan et al. (2001) show that in k nearest neighbors recommender systems, the recommendations could disclose the interaction data of the neighbors, i.e., users, whose interaction data is utilized to generate the recommendations. Similarly, Calandrino et al. (2011) inject fake users to make the recommendations more likely to disclose the neighbors' interaction data, and also, they can infer users' interaction data based on the public outputs of a recommender system, e.g., public interaction data or public product reviews. Furthermore, Hashemi et al. (2022) and Xin et al. (2023) aim to learn user behavior via observing many recommendations and, in this way, can disclose parts of a user's interaction data. Weinsberg et al. (2012) show that an adversary could infer sensitive attributes, in this case, gender, based on a user's interaction data. Their attack relies on a classifier that leverages a small set of training examples to learn the correlation between a user's preferences and gender. Likewise, Ganhör et al. (2022) show that recommender systems based on autoencoder architectures are vulnerable to infer the user's gender from the latent user representation. The authors also propose an adversarial training regime to mitigate this problem. Similarly, also Zhang et al. (2023) infer the age and gender of users in a federated learning recommender system. In summary, many of a user's sensitive attributes can be inferred via thoroughly analyzing the user's digital footprint (e.g., the behavior in a recommender system or social media platform) (Kosinski et al., 2013).

Overall, the utilization of users' interaction data for generating recommendations poses a privacy risk for users. Therefore, privacy-enhancing techniques, such as homomorphic encryption (Gentry, 2009), federated learning (McMahan et al., 2017), or most prominently, *differential privacy (DP)* (Dwork, 2008) have been applied to protect users' privacy. Specifically, DP is applied via injecting noise into the recommender system. This ensures that the recommender system uses noisy data instead of the real data. For example, an additive mechanism samples random noise from the Laplace or Gaussian distribution and adds it to the users' rating data (Dwork and Roth, 2014). Alternatively, the randomized responses mechanism flips a fair coin, which decides whether to use the real data or random data, and this way, ensures DP (Warner, 1965; Dwork and Roth, 2014). Overall, the degree of noise that is used is defined by the parameter ϵ, i.e., the privacy budget. Intuitively, the smaller the ϵ-value is, the better the privacy, but the stronger the expected accuracy drop. Therefore, choosing ϵ is non-trivial and depends on the specific use case (Dwork, 2008).

3. Review methodology

To conduct our review, we chose relevant conferences in the field, i.e., ACM SIGIR, TheWebConf, ACM KDD, IJCAI, ACM CIKM, and ACM RecSys and journals, i.e., TOIS, TIST, UMUAI, and TKDE. Adopting a keyword-based search, we identified relevant publications in the proceedings via querying the full-texts for "differential privacy" and "recommender system", "recommend", "recommendation", or "recommender". We manually checked the resulting papers for their relevance and retrieved 16 publications. In addition, we conducted a literature search on Google Scholar using the same keywords and procedure, which resulted in 10 publications. Overall, we considered 26 publications in the paper at hand.

4. Recommender systems with differential privacy

According to Friedman et al. (2016), DP can be applied via (i) adding noise to the input of a collaborative filtering-based recommender system, e.g., the user data or other user representations, (ii) adding noise to the training process of the model, i.e., the model updates, or (iii) adding noise to the model after training, i.e., to the resulting latent factors. In Table 1, we group the selected publications into these three categories.

4.1. Differential privacy applied to the user representation

In collaborative filtering recommender systems, the input to the system is typically given by interaction or rating data. However, more complex user representations exist, e.g., neural-based user embeddings.

Chen et al. (2020) protect POI (point of interest) interaction data of users, e.g., a user visited a restaurant, with DP. Specifically, they use this data to privately calculate POI features, i.e., the

TABLE 1 Overview of the reviewed 26 publications.

References	Domain(s)	DP applied to		
		User represent.	Model updates	After training
Long et al. (2023)	Location	•		
Müllner et al. (2023)	Movies, Music, Books, Social	•		
Neera et al. (2023)	Movies, Jokes, Dating	•		
Wang et al. (2023)	Movies, Music		•	
Chai et al. (2022)	Movies, Location	•		
Chen et al. (2022)	Movies, Music, Books	•		
Jiang et al. (2022)	Movies, Music, Location, Groceries		•	
Liu et al. (2022)	Social		•	
Ning et al. (2022)	Movies		•	
Ran et al. (2022)	Movies, Music			•
Ren et al. (2022)	Social	•		
Wu et al. (2022)	Advertisement	•		
Li et al. (2021)	Movies, Dating		•	
Minto et al. (2021)	Movies		•	
Zhang et al. (2021)	Movies	•		•
Chen et al. (2020)	Location	•		
Gao et al. (2020)	Movies, Smartphone	•		
Ma et al. (2019)	Health		•	
Meng et al. (2018)	Social		•	
Shin et al. (2018)	Movies, Dating		•	
Liu et al. (2017)	Movies	•		
Yang et al. (2017)	Movies	•		
Li et al. (2016)	Movies	•		
Hua et al. (2015)	Movies		•	•
Zhu et al. (2013)	Movies	•		
Zhao et al. (2011)	Movies	•		

We mark whether DP is applied to the user representation, to the model updates, or after training. Domain(s) refers to the domain(s) in which the recommendations are evaluated. We sort the publications with respect to recency.

number of visitors per restaurant, which are subsequently used for generating recommendations instead of the DP-protected interaction data. This way, they can increase recommendation accuracy. Similarly, Long et al. (2023) use DP to recommend POIs, but in a decentralized fashion. A central server collects public data to train a recommendation model and to privately identify groups of similar users. DP is used for privately calculating user-user similarities. Then, users locally use information from similar users, which leads to a better trade-off between recommendation quality and privacy than comparable approaches.

Liu et al. (2017) add noise to users' rating data and to the user-user covariance matrix to ensure DP of a KNN-based recommender system. They show that this leads to better privacy than in case only the covariance matrix is protected via DP. Besides revealing users' rating data, an attacker could also aim to infer sensitive attributes (e.g., gender) of the users. Therefore, Chai et al. (2022) propose an obfuscation model to protect gender information. After applying this obfuscation model, users protect their data via DP and send it to a central server. Yang et al. (2017) use the Johnson-Lindenstrauss transform (Blocki et al., 2012), i.e., they ensure DP via multiplying the original interaction matrix with a random matrix. Using this protected matrix, their approach guarantees differential privacy and also can even generate more accurate recommendations than a non-private approach. Neera et al. (2023) underline that adding Laplacian noise can lead to "unrealistic" rating values, i.e., outside the rating range, and through this, recommendation accuracy can drop severely. Therefore, they bound the noisy ratings to a "realistic" value range without harming DP. Plus, they use a Gaussian mixture model to estimate and then remove noise in the recommendation process to keep recommendation accuracy.

Cross-domain recommendation models can increase recommendation accuracy in the target domain by exploiting data from multiple source domains. To protect user privacy when data from the source domain is made available to the target domain, Chen et al. (2022) use the Johnson-Lindenstrauss transform. Due to the high sparsity of the rating matrix, they employ a variant that performs better when applied to sparse matrices (Ailon and Chazelle, 2009). Ren et al. (2022) utilize data from different social network platforms to generate recommendations and apply DP to the user attributes and the connections in the social network graphs. Plus, they apply a variant of DP to protect textual data (Fernandes et al., 2019). Moreover, to increase the click-through rate for recommended advertisements, Wu et al. (2022) leverage user interaction data from multiple platforms. First, user embeddings are generated per platform and then protected with DP. Second, the recommender system collects and aggregates a user's DP-protected embeddings across platforms and then applies DP again to the aggregated user embedding. According to the authors, applying DP after aggregation allows for smaller noise levels when applying DP to the per-platform user embeddings, which results in higher accuracy. Typically, many users use a variety of different online platforms. Therefore, Li et al. (2016) leverage these multiple data sources per user to increase recommendation accuracy. Specifically, they combine DP-protected item-item similarities from dataset *B* as auxiliary data that helps to generate more accurate recommendations for users in dataset *A* (cf. Zhao et al., 2011).

Gao et al. (2020) compute item-item similarities by using DP-protected user interaction data. With these item-item similarities, users can locally generate recommendations on their own devices, therefore not harming their privacy. The item-based KNN recommender system proposed by Zhu et al. (2013) utilizes DP in two ways: First, they randomly rearrange the most similar neighbors to foster privacy. Second, they measure how the item-item similarity changes if a specific user interaction was not present, and with this, they add the necessary level of noise to the users' interactions. This way, recommendation accuracy can be better preserved than with approaches that apply the same level of noise to all user interactions. For user-based KNN, Müllner et al. (2023) identify neighbors that can be reused for many recommendations. This way, only a small set of users are used as neighbors for many recommendations and need to be protected with DP. Many users, however, are only rarely utilized as neighbors and therefore do not need to be protected with DP. Overall, this yields more accurate recommendations than in case DP needs to be applied to all users.

4.2. Differential privacy applied to the model updates

Some recommender systems do not process user data and create user representations on a central server, instead, they compute the model updates, i.e., gradients, locally on their users' device. Then, the recommender system collects these gradients to adapt its recommendation model. To prohibit the leakage of user data through these gradients (Bagdasaryan et al., 2020), DP can be applied.

For example, Hua et al. (2015) add noise to the gradients of the recommendation model to ensure DP. However, due to the sparsity of the gradients, the application of DP can be ineffective and information about what items have been rated by the user can be disclosed. To address this problem, Shin et al. (2018) use DP to mask whether a user appears in the dataset. Also, they formally show that the noise added to the gradients hinders a fast convergence of the recommendation model, and in this way, increases the training time. Therefore, they introduce a stabilization factor to enable better training of the recommendation model. Wang et al. (2023) propose a recommender system that uses a special DP-mechanism (Zhao et al., 2020) to simultaneously protect the rating values and the set of items that is rated by a user. The DP-protected item-vectors are then send to a central server, which performs dimensionality reduction to reduce the accuracy drop (cf. Shin et al., 2018). In Minto et al. (2021), users receive a global model from a central server and, then, compute their respective updates locally. These updates are protected via DP, before being sent back to the server. Plus, the number of updates per user are restricted to further improve privacy. Moreover, the authors highlight that high-dimensional gradients can negatively impact the recommendation quality, as they are especially prone to higher sparsity (cf. Hua et al., 2015; Shin et al., 2018). When DP is applied, the gradients become denser since noise is added to the entire gradient, including the zero-entries. This, in practice, leads to additional communication overhead, since all non-zero-entries need to be transmitted (Ning et al., 2022). Therefore, Ning et al. only add noise to the non-zero gradients. This way, the communication overhead is reduced; however, DP cannot be guaranteed anymore.

Jiang et al. (2022) reduce the accuracy drop via an adaptive DP mechanism that depends on the number of training steps. Intuitively, after many training steps, the model fine-tunes its predictions and the gradients need to be measured more accurately than during the beginning of the model training. Thus, they add more noise in the beginning and less noise in the end of the training process. This yields more accurate recommendations than a static DP mechanism that always adds the same level of noise. Li et al. (2021) also use noisy model updates to ensure DP. They observe that noise can lead to large values for the user embeddings, which increases the sensitivity and therefore also the level of noise that is required to ensure DP. To foster recommendation quality, they map the user embeddings to a certain range, which bounds the sensitivity and requires less noise. Liu et al. (2022) leverage user interactions and social connections to generate recommendations via a federated graph neural network. To ensure DP, they add noise to the gradients that are sent to a central server. However, gradients with different magnitudes have different sensitivities (cf. Li et al., 2021), and thus, need a different level of noise to ensure DP. Therefore, they fit the noise level to the gradient magnitudes to satisfy DP, but also, to preserve recommendation accuracy.

Ma et al. (2019) employ federated tensor factorization in the health domain. A global model is distributed to hospitals, which locally update the model based on their data. To protect privacy, a variant of DP is applied to the model updates, which are subsequently sent to the global server to adapt the global model. Meng et al. (2018) randomly divide users' ratings into non-sensitive and sensitive ratings. For sensitive ratings, they apply more noise than for non-sensitive ratings. With this, their approach can preserve higher recommendation accuracy than in case the same noise level is used for sensitive and non-sensitive data.

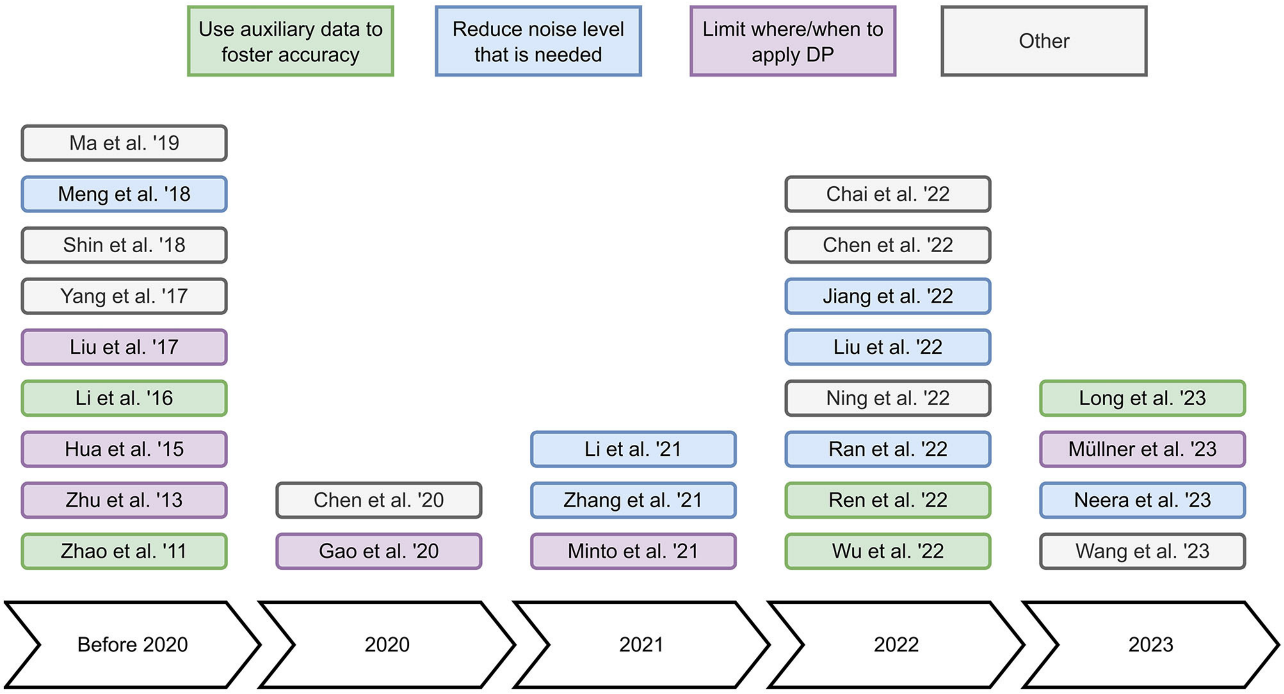

FIGURE 1
Overview of the main concepts of the reviewed papers. *Use auxiliary data to foster accuracy* refers to the incorporation of data from other domains, datasets or users to increase recommendation accuracy. *Reduce noise level that is needed* refers to designing recommender systems that require a minimal amount of noise to ensure DP. *Limit where/when to apply DP* refers to carefully minimizing the application of DP. *Other* refers to approaches that do not fit into the previous categories.

4.3. Differential privacy applied after training

Only few works apply DP to the recommendation model after training. In case of a matrix factorization approach, noise can be added to the learned user- and item-vectors to ensure DP. Our selected publications (see Section 3) do not include any works that apply DP exclusively to the model after training. Nevertheless, we describe works that apply DP to the user representation or the model updates, but also after training.

For example, Hua et al. (2015) consider a matrix factorization model, where the model sends item-vectors back to the users and this way, users' data can get leaked. To prohibit this, Hua et al. perturb the model's objective function after training via adding noise to the latent item-vectors. Similarly, Ran et al. (2022) also use DP to prohibit data leakage through the item-vectors that are sent to the users. Specifically, a trusted recommender system generates a matrix factorization model. Instead of publishing the item-vectors of this model, they learn new item-vectors on the DP-protected user-vectors. Through this, they can minimize the noise that is introduced and thus, can improve recommendation accuracy over comparable approaches. Zhang et al. (2021) apply DP to the user representation and also, to the model after training. Specifically, they use a polynomial approximation of the model's loss function to efficiently compute the sensitivity of the dataset and, accordingly, adapt the level of noise that is added to the loss function.

5. Summary and open questions

In this review, we investigate research works that apply DP to collaborative filtering recommender systems. We identify 26 relevant works and categorize these based on how they apply DP, i.e., to the user representation, to the model updates, or to the model after training (see Table 1). In addition, we briefly summarize these relevant works to obtain a broad overview of the state-of-the-art. Furthermore, we identify the main concepts of the relevant works in Figure 1 to help readers to understand in which diverse ways the reviewed papers apply DP to improve the accuracy-privacy trade-off. Our main findings from reviewing the discussed literature are two-fold: (i) The majority of works use datasets from the same non-sensitive domain, i.e., movies, and (ii) applying DP to the model after training seems to be an understudied topic.

Many research works use datasets from the movie domain, which, in general, does not include sensitive data. For research on DP in collaborative filtering recommender systems, however, datasets from sensitive domains may be better suited to resemble real-world privacy threats well. For example, datasets from the health, finance, or job domain. Moreover, the majority of research focuses on either applying DP to the user representation or to the model updates. Research on applying DP to the model after training is scarce, and therefore, this opens up the possibility of future work to fill this gap.

Our review of relevant work allows to grasp the state-of-the-art and to identify the following open research questions:

Q1: How does applying DP impact fairness? Dwork et al. (2012) and Zemel et al. (2013) suggest that in theory, privacy can lead to fairness and fairness can lead to privacy. The reason is that for both, a user's data shall be hidden, either to ensure privacy or to prohibit discrimination based on this data. However, in practice, correlations in private data can still lead to unfairness (Ekstrand et al., 2018; Agarwal, 2020). Only recently, Yang et al. (2023) and Sun et al. (2023) investigate the connection between privacy and fairness in recommender systems. For example, Sun et al. (2023) use DP-protected information to re-rank the items in the recommendation list and in this way, increase a more fair exposure of items. Nonetheless, the impact of DP on fairness remains an understudied topic.

Q2: How to quantify the user's perceived privacy? Users perceive privacy differently, e.g., some users tolerate disclosing their gender, while others refuse to do this (Joshaghani et al., 2018). This perceived privacy depends on many factors, e.g., context or situational factors (Knijnenburg and Kobsa, 2013; Mehdy et al., 2021). However, measuring users' perceived privacy is hard and is usually done via questionnaires (Knijnenburg and Kobsa, 2013). This is in stark contrast to how privacy is measured in the DP framework, i.e., via quantifying to what extent the data impacts the output of the recommender system. Therefore, developing methods to better quantify users' privacy is an important future research avenue.

Author contributions

PM: literature analysis, conceptualization, and writing. MS: conceptualization and writing. EL and DK: conceptualization, writing, and supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the DDAI COMET Module within the COMET-Competence Centers for Excellent Technologies Programme, funded by the Austrian Federal Ministry for Transport, Innovation and Technology (BMVIT), the Austrian Federal Ministry for Digital and Economic Affairs (BMDW), the Austrian Research Promotion Agency (FFG), the province of Styria (SFG), and partners from industry and academia. The COMET Programme is managed by FFG. In addition, the work received funding from the TU Graz Open Access Publishing Fund and from the Austrian Science Fund (FWF): DFH-23 and P33526.

Conflict of interest

PM was employed by Know-Center Gmbh.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Agarwal, S. (2020). *Trade-offs between fairness, interpretability, and privacy in machine learning* (Master's thesis). University of Waterloo, Waterloo, ON, Canada.

Ailon, N., and Chazelle, B. (2009). The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.* 39, 302–322. doi: 10.1137/060673096

Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., and Shmatikov, V. (2020). "How to backdoor federated learning," in *International Conference on Artificial Intelligence and Statistics* (Palermo: PMLR), 2938–2948.

Beigi, G., and Liu, H. (2020). A survey on privacy in social media: identification, mitigation, and applications. *ACM Trans. Data Sci.* 1, 1–38. doi: 10.1145/3343038

Berkovsky, S., Kuflik, T., and Ricci, F. (2012). The impact of data obfuscation on the accuracy of collaborative filtering. *Expert Syst. Appl.* 39, 5033–5042. doi: 10.1016/j.eswa.2011.11.037

Bilge, A., Kaleli, C., Yakut, I., Gunes, I., and Polat, H. (2013). A survey of privacy-preserving collaborative filtering schemes. *Int. J. Softw. Eng. Knowledge Eng.* 23, 1085–1108. doi: 10.1142/S0218194013500320

Blocki, J., Blum, A., Datta, A., and Sheffet, O. (2012). "The Johnson-Lindenstrauss transform itself preserves differential privacy," in *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science* (New Brunswick, NJ: IEEE), 410–419.

Calandrino, J. A., Kilzer, A., Narayanan, A., Felten, E. W., and Shmatikov, V. (2011). ""You might also like:" privacy risks of collaborative filtering," in *Proc. of S&P'11* (Oakland, CA: IEEE), 231–246.

Chai, D., Wang, L., Chen, K., and Yang, Q. (2022). Efficient federated matrix factorization against inference attacks. *ACM Trans. Intell. Syst. Technol.* 13, 1–20. doi: 10.1145/3501812

Chen, C., Wu, H., Su, J., Lyu, L., Zheng, X., and Wang, L. (2022). "Differential private knowledge transfer for privacy-preserving cross-domain recommendation," in *Proc. of ACM WWW'22* (Lyon).

Chen, C., Zhou, J., Wu, B., Fang, W., Wang, L., Qi, Y., et al. (2020). Practical privacy preserving POI recommendation. *ACM Trans. Intell. Syst. Technol.* 11, 1455–1465. doi: 10.1145/3394138

Dwork, C. (2008). "Differential privacy: a survey of results," in *International Conference on Theory and Applications of Models of Computation* (Berlin: Springer), 1–19.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. (2012). "Fairness through awareness," in *Proc. of ITCS'12* (Cambridge, MA), 214–226.

Dwork, C., and Roth, A. (2014). The algorithmic foundations of differential privacy. *Found. Trends Theoret. Comput. Sci.* 9, 211–407. doi: 10.1561/0400000042

Ekstrand, M. D., Joshaghani, R., and Mehrpouyan, H. (2018). "Privacy for all: Ensuring fair and equitable privacy protections," in *Proc. of FAccT'18* (New York, NY: PMLR), 35–47.

Fernandes, N., Dras, M., and McIver, A. (2019). "Generalised differential privacy for text document processing," in *Principles of Security and Trust: 8th International Conference, POST 2019* (Prague: Springer International Publishing), 123–148.

Friedman, A., Berkovsky, S., and Kaafar, M. A. (2016). A differential privacy framework for matrix factorization recommender systems. *User Model. User Adapt. Interact.* 26, 425–458. doi: 10.1007/s11257-016-9177-7

Friedman, A., Knijnenburg, B. P., Vanhecke, K., Martens, L., and Berkovsky, S. (2015). "Privacy aspects of recommender systems," in *Recommender Systems Handbook*, eds F. Ricci, L. Rokach, and B. Shapira (Boston, MA: Springer), 649–688.

Ganhör, C., Penz, D., Rekabsaz, N., Lesota, O., and Schedl, M. (2022). "Unlearning protected user attributes in recommendations with adversarial training," in *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, eds E. Amigó, P. Castells, J. Gonzalo, B. Carterette, J. S. Culpepper, and G. Kazai (Madrid: ACM), 2142–2147.

Gao, C., Huang, C., Lin, D., Jin, D., and Li, Y. (2020). "DPLCF: differentially private local collaborative filtering," in *Proc. of SIGIR'20* (Xi'an), 961–970.

Gentry, C. (2009). *A Fully Homomorphic Encryption Scheme (Dissertation)*. Stanford University.

Hashemi, H., Xiong, W., Ke, L., Maeng, K., Annavaram, M., Suh, G. E., et al. (2022). Data leakage via access patterns of sparse features in deep learning-based recommendation systems. *arXiv preprint arXiv:2212.06264*.

Herbert, C., Marschin, V., Erb, B., Meißner, D., Aufheimer, M., and Bösch, C. (2021). Are you willing to self-disclose for science? Effects of privacy awareness and trust in privacy on self-disclosure of personal and health data in online scientific studies–an experimental study. *Front. Big Data* 4, 763196. doi: 10.3389/fdata.2021.763196

Himeur, Y., Sohail, S. S., Bensaali, F., Amira, A., and Alazab, M. (2022). Latest trends of security and privacy in recommender systems: a comprehensive review and future perspectives. *Comput. Sec.* 118, 102746. doi: 10.1016/j.cose.2022.102746

Hua, J., Xia, C., and Zhong, S. (2015). "Differentially private matrix factorization," in *International Joint Conference on Artificial Intelligence* (Buenos Aires).

Jeckmans, A. J., Beye, M., Erkin, Z., Hartel, P., Lagendijk, R. L., and Tang, Q. (2013). "Privacy in recommender systems," in *Social Media Retrieval*, eds N. Ramzan, R. van Zwol, J.-S. Lee, K. Clüver, and X.-S. Hua (London: Springer), 263–281.

Jiang, X., Liu, B., Qin, J., Zhang, Y., and Qian, J. (2022). "FedNCF: federated neural collaborative filtering for privacy-preserving recommender system," in *2022 International Joint Conference on Neural Networks (IJCNN)* (Padua: IEEE), 1–8.

Joshaghani, R., Ekstrand, M. D., Knijnenburg, B., and Mehrpouyan, H. (2018). "Do different groups have comparable privacy tradeoffs?" in *Workshop on Moving Beyond a "One-Size Fits All" Approach: Exploring Individual Differences in Privacy, in Conjunction with the ACM CHI Conference on Human Factors in Computing Systems, CHI 2018* (Montreal, QC).

Knijnenburg, B. P., and Kobsa, A. (2013). Making decisions about privacy: information disclosure in context-aware recommender systems. *ACM Trans. Interact. Intell. Syst.* 3, 1–23. doi: 10.1145/2499670

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Nat. Acad. Sci. U.S.A.* 110, 5802–5805. doi: 10.1073/pnas.1218772110

Li, J., Yang, J.-J., Zhao, Y., Liu, B., Zhou, M., Bi, J., and Wang, Q. (2016). Enforcing differential privacy for shared collaborative filtering. *IEEE Access* 5, 35–49. doi: 10.1109/ACCESS.2016.2600258

Li, Z., Ding, B., Zhang, C., Li, N., and Zhou, J. (2021). Federated matrix factorization with privacy guarantee. *Proc. VLDB Endowment* 15, 900–913. doi: 10.14778/3503585.3503598

Liu, X., Liu, A., Zhang, X., Li, Z., Liu, G., Zhao, L., and Zhou, X. (2017). "When differential privacy meets randomized perturbation: a hybrid approach for privacy-preserving recommender system," in *International Conference on Database Systems for Advanced Applications* (Suzhou: Springer), 576–591.

Liu, Z., Yang, L., Fan, Z., Peng, H., and Yu, P. S. (2022). Federated social recommendation with graph neural network. *ACM Trans. Intell. Syst. Technol.* 13, 1–24. doi: 10.1145/3501815

Long, J., Chen, T., Nguyen, Q. V. H., and Yin, H. (2023). Decentralized collaborative learning framework for next POI recommendation. *ACM Trans. Inf. Syst.* 41, 1–25. doi: 10.1145/3555374

Ma, J., Zhang, Q., Lou, J., Ho, J. C., Xiong, L., and Jiang, X. (2019). "Privacy-preserving tensor factorization for collaborative health data analysis," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19* (New York, NY: Association for Computing Machinery), 1291–1300.

Majeed, A., and Lee, S. (2020). Anonymization techniques for privacy preserving data publishing: a comprehensive survey. *IEEE Access* 9, 8512–8545. doi: 10.1109/ACCESS.2020.3045700

McMahan, B., Moore, E., Ramage, D., Hampson, S., and Arcas, B. A. (2017). "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics* (Fort Lauderdale, FL: PMLR), 1273–1282.

Mehdy, A. N., Ekstrand, M. D., Knijnenburg, B. P., and Mehrpouyan, H. (2021). "Privacy as a planned behavior: effects of situational factors on privacy perceptions and plans," in *Proc. of UMAP'21* (Utrecht).

Meng, X., Wang, S., Shu, K., Li, J., Chen, B., Liu, H., et al. (2018). "Personalized privacy-preserving social recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence* (New Orleans, LA).

Minto, L., Haller, M., Livshits, B., and Haddadi, H. (2021). "Stronger privacy for federated collaborative filtering with implicit feedback," in *Proceedings of the 15th ACM Conference on Recommender Systems* (Amsterdam), 342–350.

Müllner, P., Lex, E., Schedl, M., and Kowald, D. (2023). ReuseKNN: neighborhood reuse for differentially-private KNN-based recommendations. *ACM Trans. Intell. Syst. Technol.* 14, 1–29. doi: 10.1145/3608481

Neera, J., Chen, X., Aslam, N., Wang, K., and Shu, Z. (2023). Private and utility enhanced recommendations with local differential privacy and Gaussian mixture model. *IEEE Trans. Knowledge Data Eng.* 35, 4151–4163. doi: 10.1109/TKDE.2021.3126577

Ning, L., Chien, S., Song, S., Chen, M., Xue, Y., and Berlowitz, D. (2022). "EANA: reducing privacy risk on large-scale recommendation models," in *Proceedings of the 16th ACM Conference on Recommender Systems* (Seattle, WA), 399–407.

Ramakrishnan, N., Keller, B. J., Mirza, B. J., Grama, A. Y., and Karypis, G. (2001). When being weak is brave: privacy in recommender systems. *IEEE Internet Comput.* 5, 54–62. doi: 10.1109/4236.968832

Ran, X., Wang, Y., Zhang, L. Y., and Ma, J. (2022). A differentially private matrix factorization based on vector perturbation for recommender system. *Neurocomputing* 483, 32–41. doi: 10.1016/j.neucom.2022.01.079

Ren, J., Jiang, L., Peng, H., Lyu, L., Liu, Z., Chen, C., et al. (2022). "Cross-network social user embedding with hybrid differential privacy guarantees," in *Proceedings of the 31st ACM International Conference on Information &; Knowledge Management, CIKM '22* (New York, NY: Association for Computing Machinery), 1685–1695.

Shin, H., Kim, S., Shin, J., and Xiao, X. (2018). Privacy enhanced matrix factorization for recommendation with local differential privacy. *IEEE Trans. Knowledge Data Eng.* 30, 1770–1782. doi: 10.1109/TKDE.2018.2805356

Sun, J. A., Pentyala, S., Cock, M. D., and Farnadi, G. (2023). "Privacy-preserving fair item ranking," in *Advances in Information Retrieval*, eds J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, and A. Caputo (Cham: Springer Nature), 188–203.

Wang, Y., Gao, M., Ran, X., Ma, J., and Zhang, L. Y. (2023). An improved matrix factorization with local differential privacy based on piecewise mechanism for recommendation systems. *Expert Syst. Appl.* 216, 119457. doi: 10.1016/j.eswa.2022.119457

Warner, S. L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *J. Am. Stat. Assoc.* 60, 63–69.

Weinsberg, U., Bhagat, S., Ioannidis, S., and Taft, N. (2012). "Blurme: inferring and obfuscating user gender based on ratings," in *Proceedings of the Sixth ACM Conference on Recommender Systems* (Dublin), 195–202.

Wu, C., Wu, F., Lyu, L., Huang, Y., and Xie, X. (2022). FedCTR: Federated native ad CTR prediction with cross-platform user behavior data. *ACM TIST* 13, 1–19. doi: 10.1145/3506715

Xin, X., Yang, J., Wang, H., Ma, J., Ren, P., Luo, H., et al. (2023). On the user behavior leakage from recommender system exposure. *ACM Trans. Inform. Syst.* 41, 1–25. doi: 10.1145/3568954

Yang, M., Zhu, T., Ma, L., Xiang, Y., and Zhou, W. (2017). "Privacy preserving collaborative filtering via the Johnson-Lindenstrauss transform," in *2017 IEEE Trustcom/BigDataSE/ICESS* (Sydney, NSW: IEEE), 417–424.

Yang, Z., Ge, Y., Su, C., Wang, D., Zhao, X., and Ying, Y. (2023). "Fairness-aware differentially private collaborative filtering," in *Companion Proceedings of the ACM Web Conference 2023, WWW '23 Companion* (Austin, TX: Association for Computing Machinery), 927–931.

Zemel, R., Wu, Y., Swersky, K., Pitassi, T., and Dwork, C. (2013). "Learning fair representations," in *Proc. of ICML'13* (Atlanta, GA: PMLR), 325–333.

Zhang, S., Yin, H., Chen, T., Huang, Z., Cui, L., and Zhang, X. (2021). "Graph embedding for recommendation against attribute inference attacks," in *Proceedings of the Web Conference 2021, WWW '21* (New York, NY: Association for Computing Machinery), 3002–3014.

Zhang, S., Yuan, W., and Yin, H. (2023). Comprehensive privacy analysis on federated recommender system against attribute inference attacks. *IEEE Trans. Knowledge Data Eng.* 1-13. doi: 10.1109/TKDE.2023.3295601

Zhao, Y., Feng, X., Li, J., and Liu, B. (2011). "Shared collaborative filtering," in *Proceedings of the Fifth ACM Conference on Recommender Systems* (Chicago, IL), 29–36.

Zhao, Y., Zhao, J., Yang, M., Wang, T., Wang, N., Lyu, L., et al. (2020). Local differential privacy-based federated learning for internet of things. *IEEE Internet Things J.* 8, 8836–8853. doi: 10.1109/JIOT.2020.30 37194

Zhu, T., Li, G., Ren, Y., Zhou, W., and Xiong, P. (2013). "Differential privacy for neighborhood-based collaborative filtering," in *Proc. of IEEE/ACM ASONAM'13* (Niagara Falls, ON), 752–759.

frontiers | Frontiers in Big Data

TYPE Review
PUBLISHED 30 October 2023
DOI 10.3389/fdata.2023.1284511

Check for updates

OPEN ACCESS

EDITED BY
Dominik Kowald,
Know Center, Austria

REVIEWED BY
Tomislav Duricic,
Graz University of Technology, Austria
Simone Kopeinik,
Know Center, Austria
Peter Müllner,
Know Center Graz, Austria, in collaboration with reviewer SK

*CORRESPONDENCE
Alexander Felfernig
alexander.felfernig@ist.tugraz.at

RECEIVED 28 August 2023
ACCEPTED 09 October 2023
PUBLISHED 30 October 2023

CITATION
Felfernig A, Wundara M, Tran TNT, Polat-Erdeniz S, Lubos S, El Mansi M, Garber D and Le V-M (2023) Recommender systems for sustainability: overview and research issues. *Front. Big Data* 6:1284511. doi: 10.3389/fdata.2023.1284511

COPYRIGHT
© 2023 Felfernig, Wundara, Tran, Polat-Erdeniz, Lubos, El Mansi, Garber and Le. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Recommender systems for sustainability: overview and research issues

Alexander Felfernig[1]*, Manfred Wundara[2], Thi Ngoc Trang Tran[1], Seda Polat-Erdeniz[1], Sebastian Lubos[1], Merfat El Mansi[1], Damian Garber[1] and Viet-Man Le[1]

[1]Applied Software Engineering & AI, Institute of Software Technology, Graz University of Technology, Graz, Austria, [2]Magistrat Villach, Villach, Austria

Sustainability development goals (SDGs) are regarded as a universal call to action with the overall objectives of planet protection, ending of poverty, and ensuring peace and prosperity for all people. In order to achieve these objectives, different AI technologies play a major role. Specifically, recommender systems can provide support for organizations and individuals to achieve the defined goals. Recommender systems integrate AI technologies such as machine learning, explainable AI (XAI), case-based reasoning, and constraint solving in order to find and explain user-relevant alternatives from a potentially large set of options. In this article, we summarize the state of the art in applying recommender systems to support the achievement of sustainability development goals. In this context, we discuss open issues for future research.

KEYWORDS

sustainability, recommender systems, machine learning, sustainability development goals, artificial intelligence

1. Introduction

The overall objective of the 17 sustainability development goals (SDGs—see Table 1; e.g., *no poverty* and *quality education*) is to provide a universal call to *end poverty*, *planet protection*, and to *ensure that people enjoy peace and prosperity* also with the goal to establish a balance of social, economic, and environmental sustainability.[1] Existing research (vanWynsberghe, 2021) has already shown that Artificial Intelligence (AI) methods and techniques can have positive as well as negative impacts ranging from efficient energy production and distribution to negative aspects such as increasing power consumption scenarios due to different types of large-scale machine learning efforts (Vinuesa et al., 2020). In this article, we analyze potentials of recommender systems as a key technology to support the mentioned SDGs.

Recommender systems can be regarded as decision support systems combining AI technologies such as machine learning, explanations, and intelligent user interfaces with the overall goal to improve a user's decision quality (Bui, 2000; Falkner et al., 2011). There are different types of recommender systems with differing applicability depending on the underlying recommendation scenario. (1) *Collaborative filtering* (CF; Ekstrand et al., 2011) follows the idea of word-of-mouth promotion where opinions of family members and friends (the so-called "nearest neighbors") are regarded as relevant recommendations for a person. (2) *Content-based Filtering* (CBF; Pazzani and Billsus, 2007) is based on

1 https://www.undp.org/sustainable-development-goals

TABLE 1 An overview of the United Nations Sustainable Development Goals (SDGs 1–17).

ID	SDG	Description
1	No poverty	End poverty in all its forms everywhere
2	Zero hunger	End hunger, achieve food security and improved nutrition, and promote sustainable agriculture
3	Good health and wellbeing	Ensure healthy lives and promote wellbeing for all at all ages
4	Quality education	Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all
5	Gender equality	Achieve gender equality and empower all women and girls
6	Clean water and sanitation	Ensure availability and sustainable management of water and sanitation for all
7	Affordable and clean energy	Ensure access to affordable, reliable, sustainable, and modern energy for all
8	Decent work and economic growth	Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all
9	Industry, innovation, and infrastructure	Build resilient infrastructure, promote inclusive, and sustainable industrialization and foster innovation
10	Reduced inequalities	Reduce inequality within and among countries
11	Sustainable cities and communities	Make cities and human settlements inclusive, safe, resilient, and sustainable
12	Responsible consumption and production	Ensure sustainable consumption and production patterns
13	Climate action	Take urgent action to combat climate change and its impacts
14	Life below water	Conserve and sustainably use the oceans, seas, and marine resources for sustainable development
15	Life on land	Protect, restore, and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss
16	Peace, justice, and strong institutions	Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable, and inclusive institutions at all levels
17	Partnerships for the goals	Strengthen the means of implementation and revitalize the global partnership for sustainable development

the idea that if a person had specific preferences in the (near) past, these preferences would more or less remain stable and can be used for future item recommendations. (3) *Knowledge-based recommender systems* (KBR; Burke, 2000) are based on the idea of determining recommendations on the basis of a more in-depth semantic knowledge expressed, for example, in terms of constraints (Felfernig and Burke, 2008) or with attribute-level similarity metrics (Chen and Pu, 2012). (4) *Hybrid recommender systems* (HYB; Burke, 2002) focus on exploiting synergy effects by trying to combine the advantages of different recommendation approaches, for example, combining CF and CBF helps to tackle the challenges of ramp-up problems (when, e.g., CF rating data are not available for a specific user). (5) *Group recommender systems* (GRP; Felfernig et al., 2018) focus on the determination of recommendations for groups, i.e., not individual users. Such approaches have to identify recommendations that help to achieve—in one way or another—a consensus among group members.[2]

In this article, we focus on indicating in which ways recommender systems can be applied to better achieve the mentioned SDGs. With this, the major contributions of our article are the following: (1) we provide an overview of the current state-of-the-art in applying recommender systems for achieving the 17 SDGs. (2) on the basis of this overview, we discuss different open issues for future research. (3) For the given SDGs, we provide concrete working examples of how to apply recommender systems. The contributions of this article enhance existing topic-related overviews (Bui, 2000; Vinuesa et al., 2020; vanWynsberghe, 2021) in terms of (1) a focus on recommender systems technologies for sustainability, (2) the provision of concrete examples of how recommender systems can be applied to achieve individual SDGs, and (3) a discussion of recommender systems specific open research issues.

Basic insights from this overview can be summarized as follows. (1) recommender systems can already be regarded as an important technology to support the achievement of sustainability development goals. For each of the existing SDGs, corresponding recommender approaches could be identified. (2) although an application majority of CF recommenders could be observed, all of mentioned recommendation approaches (CF, CBF, KBR, HYB, and GRP) have sustainability-related applications. (3) for the discussed recommender applications, two different levels of recommender "users" exist: first, a *macro-level* with more abstract organizations (e.g., countries) and second, a *micro-level* with concrete entities (e.g., citizens).

The remainder of this article is organized as follows. In Section 2, we present our methodological approach to analyze and summarize the existing state of the art in applying recommender systems to achieve sustainability development goals (SDGs). Section 3 provides an overview of the 17 SDGs and a detailed overview of the current state of the art in applying recommender systems for achieving these goals. From this discussion of the existing best-practices, we summarize related open issues for future

2 Further details on technical backgrounds of these recommendation approaches will be provided in examples introduced in Section 3.

research (see Section 4). Finally, this article is concluded within Section 5.

2. Methodology

In this article, we focus on a comprehensive overview of the existing state of the art in *recommender systems for sustainability*. Based on the gained insights, we discuss application potentials and related open issues for future research. Our analysis of the state of the art is based on a literature review with the related phases of selecting potentially relevant papers, reviewing those papers, and a discussion of the identified papers with regard to relevance for this overview article. Paper identification is based on querying existing leading research platforms with topic-related keywords. Thereafter, the identified papers have been classified with regard to their inclusion in this overview article. In this context, queries have been performed on (1) the research platforms Google Scholar,[3] ResearchGate,[4] ScienceDirect,[5] SpringerLink,[6] Elsevier,[7] IEEE,[8], and ACM[9] and (2) recommender systems related conferences and journals including ACM Recommender Systems (ACM RecSys), ACM User Modeling and User-Adapted Interaction (ACM UMAP), ACM Intelligent User Interfaces (ACM IUI), and ACM SIGCAS/SIGCHI Computing and Sustainable Societies (COMPASS). In this context, we used the initial search queries (and different combinations thereof) of "recommender systems" + "sustainability" + "sustainability goals" + "artificial intelligence" + "decision support." Using the snowballing technique (Wohlin, 2014), we analyzed further topic-relevant references starting with the original set of identified papers. Overall, we have identified 122 relevant papers which served as a basis for writing this overview.

3. Recommender systems for sustainability

In contrast to existing approaches to evaluate the impact of recommender systems which are primarily focused on different e-commerce scenarios (Jannach and Jugovac, 2019), we focus on the impact of recommender systems in terms of achieving sustainability development goals—Table 1 provides a short overview of the 17 United Nations (UN) sustainability development goals. In the following discussions, we differentiate between (1) a *macro-level* representing recommendations determined for abstract organizations (e.g., countries, company types, and types of study programmes) and (2) a *micro-level* representing recommendations determined for concrete entities (e.g., citizens, companies, and tourists). We exemplify the application of recommender systems with a focus on basic recommendation approaches, i.e., the goal in this article is to discuss application scenarios but not primarily detailed algorithmic approaches.

3 https://scholar.google.com/
4 https://www.researchgate.net/
5 https://www.sciencedirect.com/
6 https://link.springer.com/
7 https://www.elsevier.com/
8 https://www.ieee.org/
9 https://www.acm.org/

TABLE 2 Example: applying collaborative filtering for recommending advantageous items (products).

Item (product)	$Country_1$	$Country_2$	...	$Country_n$
Computer	1.5	2.2	...	1.5
Tourism	1.1	2.8	...	1.1
Wine	1.3	0.5	...	1.2
...	...	...	...	...
Automotive	3.1	2.2	...	4.1
Solar equipment	?	?	...	5.1

? indicates that a recommendation is needed.

3.1. No poverty

The related major goal is to *end poverty everywhere*. Poverty has a multitude of definitions and can be characterized in a monetary dimension in terms of not having enough money to maintain his/her livelihood—a related overview of AI methods to estimate the degree of poverty in a region/country can be found in Usmanova et al. (2022). Examples of data sources used in such contexts are, for example, household data (e.g., demographics, education, and food consumption), food price data, and e-commerce data (Usmanova et al., 2022). Poverty prediction has to be accompanied with approaches that help to counteract poverty. For example, Che (2020) show how recommendation techniques can be applied to identify export diversification strategies in such a way that a country has a latent competitive advantage (when following this strategy).

An important measure in this context is the so-called *Revealed Comparative Advantage (RCA)* score (for a country θ and product π; see Formula 1; Balassa and Noland, 1989) which is used to determine the importance of individual items (products) in the export basket of a country. In this context, $E_{\theta\pi}$ is the export value of item (product) π for country θ.

$$RCAScore_{\theta\pi} = \frac{E_{\theta\pi} \,/\, \Sigma_\pi E_{\theta\pi}}{\Sigma_\theta E_{\theta\pi} \,/\, \Sigma_\theta \Sigma_\pi E_{\theta\pi}} \tag{1}$$

In the line of Che (2020), recommendation services can be provided on the basis of the *RCAScore* of individual items. When applying collaborative filtering (CF), an item × *RCAScore* matrix summarizes the scores of items already exported by individual countries. CF can now be applied to predict the relevance (*RCAScore*) of new items not exported by individual countries up to now. In the example shown in Table 2, basic *RCAScore* information is already available for products such as *computer*, *tourism*, and *wine*.

Some countries do not export some of the products and we would like to know for which additional products (items) it would be good for a country to extend its assortment. In Table 2, "?" indicates that a recommendation is needed, for example, for $country_1$, it would be good to focus on producing and exporting

TABLE 3 Example: simplified portfolio elements (with costs per month).

Attributes	Car			House		Workers		Holidays		Food	
Domains	BMW	Renault	None	Large	Medium	1	2	Yes	No	Flexible	Restricted
Costs/income	500	350	0	2.5 k	1.5 k	2 k	3 k	150	0	600	200

solar equipment. Based on the idea of CF, the nearest neighbor of country$_1$ is country$_n$ (the nearest neighbor is regarded as a country with a similar *RCAScore* distribution) with a high relevance of exporting solar equipment. In this simplified scenario, engaging in exporting solar equipment can be regarded also as a good idea for country$_1$. For a detailed discussion of applying different CF algorithms in such application contexts, we refer to Che (2020). Furthermore, Liao et al. (2018) discuss approaches to product diversification based on the concepts of social network analysis where relationships between countries and their products are analyzed for recommendation purposes.

On the level of individuals, poverty can be triggered by various factors such as wrong investment decisions (e.g., purchasing a too expensive car and dealing with the consequences), wrong choice of personal education and employment (e.g., to stop visiting school with the consequences of problems in finding a job), and issues in handling the personal financial situation (e.g., women focusing on childcare and without a corresponding financial provision). In the following, we provide a simple example of applying a knowledge-based recommendation (Felfernig et al., 2006) approach as a basic support in investment decisions (Fano and Kurth, 2003). Table 3 provides an overview of different portfolio elements that could be selected by the user of a recommender system.

A major criterion in portfolio recommendation is that the overall *consumed resources* (*car*, *house*, *holidays*, and *food* representing, e.g., family dinner etc.) must not exceed the provided resources (income provided by *workers* per year). This resource limitation can be expressed as shown in Formula 2 where the property *workers.income* represents the monthly income of the family.

$$12 \times (car.costs + house.costs + holidays.costs + food.costs) \leq 12 \times workers.income \quad (2)$$

On the basis of such a scenario, the user of a recommender system can choose different options, for example, an expensive car and an expensive house, and immediately understand the consequences of such decisions. For example, with the current yearly income, it is impossible to have both, an expensive car and a large house. Furthermore, there also exists a scenario (portfolio) where one worker would in principle be enough to cover all of the estimated costs. Table 4 shows the extreme cases of a portfolio with *maximum costs* p.a. (45 k) and the other extreme of *minimum costs* p.a. (20.4 k).

The presented example is a simplified variant of a knowledge-based recommender system focusing on showing to the user the impacts of specific investment decisions. In situations where the defined user preferences do not allow the recommendation of a portfolio, corresponding diagnosis techniques can help to indicate

TABLE 4 Example portfolios and associated costs p.a.

Portfolio	Car	House	Holidays	Food	Total costs p.a.
Max	BMW	Large	Yes	Flexible	45 k
Min	None	Medium	No	Restricted	20.4 k

minimal changes in the users preferences in such a way that a solution can be identified.[10]

3.2. Zero hunger

The related goal is to *end hunger and to achieve improved nutrition and food security while at the same time promoting sustainable agriculture.* In contrast to the application of recommender systems in the context of healthy living (Tran et al., 2018a), a major focus of sustainability in the context of achieving *zero hunger* is to foster more conscious food consumption and to support food production processes with a clear sustainability focus (Gill et al., 2021; Bouni et al., 2022; Martini et al., 2022). A related crop diversification (recommendation), i.e., choosing and diversifying crops, can help governments to grow more crops in ones own country and with this to reduce dependencies to other countries (Gill et al., 2021). This also includes mechanisms to effectively detect crop diseases (Omara et al., 2023).

The appropriate determination of crop factors such as maturity date, soil suitability, and pesticide requirements becomes increasingly important. Not least, to be able to choose the optimal crop in the long run as well as to optimize production and to minimize additional efforts in terms of pesticides and soil fertilization. A simplified example of a potential application of recommender systems in crop selection is shown in Table 5. In this example, the question is if *crop*$_2$ (the *current* entry) could be relevant for region *D* (no corresponding experience data available). Since average temperature and soil moisture are quite similar to region *C* (the nearest neighbor—id = 5), the expected crop$_2$ output for this region is about 83% with a recommended pesticide usage p_3. In real-world settings, further parameters are needed for determining high-quality recommendations (Gill et al., 2021).

Food rescue organizations focus on collecting and delivering food donations to those in need (Shi et al., 2021). In many cases, collected food is in temporary storage at the rescue organization where it is offered to persons in need. Collecting the food from various local food providers is a logistic problem in the sense that volunteers need to be identified who are willing to take over a

10 For further related details, we refer to Felfernig et al. (2006).

TABLE 5 Example of knowledge-based (case-based) crop recommendation.

ID	Name	Region	Pesticides	Avg. temperature (cel.)	Soil moisture (%)	Output (%)
1	$Crop_1$	A	p_1	20	30	70
2	$Crop_1$	B	p_2	22	25	75
3	$Crop_2$	E	p_2	22	25	75
4	$Crop_1$	A	p_2	20	27	76
5	$Crop_2$	C	p_3	20	27	83
Current	$Crop_2$	D	?	20	27	?

TABLE 6 Example of volunteer (user) recommendation with content-based filtering. Each table row represents a (simplified) user profile, for example, the entry *drinks* = *yes* of $user_1$ indicates that $user_1$ prefers collection tasks with beverages included.

User	Region	Beverages	Meat	Bread	Vegetables
$user_1$	A	Yes	No	Yes	No
$user_2$	B	No	No	No	Yes
$user_3$	A	Yes	Yes	Yes	No
$task_{new}$	A	Yes	Yes	No	No

specific pick-up and food delivery task. Shi et al. (2021) present a recommender system that helps to identify candidate persons with a high probability of willing to perform a new collection and delivery task.

A simplified example of supporting such scenarios on the basis of content-based filtering is depicted in Table 6. In this setting, a new collection task is defined for region *A* and includes beverages and meat. Important to know is that many food rescue organizations allow their volunteers to claim a low share of each cartload for their own. Based on this assumption, a content-based recommender system can identify those potential drivers (volunteers) who might be interested in performing the collection task. In our example, $user_3$ can be regarded as having preferences which are most similar to those of $task_{new}$—consequently, $user_3$ can be regarded as the first candidate to be contacted.

For sure, in real-world settings, further related parameters can play an important role in recommending volunteers. Examples of such parameters are *availability* (a user might be available only during specific time periods), *fairness* (all volunteers should have near-equal chances to be contacted), and *reliability* (e.g., the driver always in-time). A detailed discussion of the application of recommender systems in a food rescue scenario is given in Shi et al. (2021).

3.3. Good health and wellbeing

The related goal is *to ensure healthy lives and promote wellbeing*. The success of public health campaigns heavily depends on the appropriateness of health messages delivered to users (Cappella et al., 2015). In such scenarios, recommender systems can help to personalize message delivery given some knowledge about features and topics of interest for a user. A simple approach can be a topic-wise recommendation where new messages/campaigns are forwarded to citizens in a personalized fashion. A related simplified example is depicted in Table 7: user interests are stored in a corresponding user profile, for example, $user_3$ has a high interest in healthy eating and healthy cooking. A new health campaign should be issued and the task is to identify those users with some basic potential interest in the related topics. The most relevant topics of $message_{new}$ are *healthy eating* and *healthy cooking*—in this scenario $user_3$ and to some extent $user_2$ have related interests, i.e., these users should be contacted in the context of the new campaign. As such, this is a simple example of applying content-based filtering in the context of delivering public health campaigns (Cappella et al., 2015). To assure that users get also in touch with new topics, diversity-enhanced and collaborative recommendation can be applied to increase serendipity effects (Ravanmehr, 2021).

Another related example on the macro-level is the support of machine learning and recommender systems in the context of vaccine allocation and distribution where appropriate planning and fairness aspects play a major role (Blasioli et al., 2023). In this scenario, aspects such as population size, percentage of individuals who have already received a previous dose, and storage capacity for the vaccines are important factors to be taken into account. An overview of the application of recommender systems in the healthcare domain is provided, for example, in Tran et al. (2018b). Important to mention, related applications are quite diverse and not all of those can be discussed in this article. Examples of recommender systems in the healthcare domain range from healthy food recommendation (Wang et al., 2021), personal wellbeing (Arévalo et al., 2022), air pollution aware outdoor activity recommendation (Alcaraz-Herrera et al., 2022), context-aware sleep health recommenders (Liang, 2022), context-aware recommenders for diabetes patients (Abu-Issa et al., 2023), activity recommenders for elderly (Herpich et al., 2017), to the recommendation of healthcare professionals (Singh et al., 2023).

A simplified example of an approach to recommend food items in a healthiness-aware fashion (and—at the same time—to take into account food preferences of the current user) is apply collaborative filtering for selecting food items and then to filter relevant items using a knowledge-based approach. Table 8 depicts a collection of recipes (for simplicity, we assume main dishes) and corresponding user preferences. The *current* user has already consumed *schnitzel* and *lasagne* in the past. A recommender

TABLE 7 Example personalized message delivery in public health campaigns.

User	Healthy eating	Athletic sports	Endurance sports	Healthy cooking	Sports events
$user_1$	0.5	0.8	0.0	0.2	0.0
$user_2$	0.5	0.1	0.6	0.5	0.8
$user_3$	0.9	0.2	0.5	0.9	0.2
$message_{new}$	0.9	0.1	0.5	0.9	0.0

TABLE 8 Example food item consumption with corresponding front-of-pack labels (*a* ..*e*) where *a* indicates high and *b* low nutritional values (Julia et al., 2021).

User	$Schnitzel_e$	$Beans_a$	$Soja_b$	$Veal_d$	$Lasagne_c$	$Trout_b$	$Spaghetti_b$	$Spinach_a$	$Salad_a$
$user_1$	x			x	x			x	x
$user_2$		x	x			x		x	x
$user_3$			x						x
current	x	?	?	?	x	?	?	?	?

TABLE 9 Example group decision setting regarding the establishment of a new study program, for example, Artificial Intelligence (AI). Individual stakeholders s_i give feedback on individual proposals in terms of evaluating the interest dimensions (F)easibility and (I)nterest.

Stakeholder	AI		AI and decision making		Data science		AI in software	
	F	I	F	I	F	I	F	I
s_1	8	6	6	8	8	4	8	8
s_2	10	9	2	4	8	2	8	8
s_3	7	7	8	8	4	2	8	9
s_4	10	10	4	7	3	3	6	7
Avg	8.75	8	5	6.75	5.75	2.75	7.5	8

could recommend these or similar items also in the future (e.g., *veal*). However, since both selections have rather low nutritional values (Julia et al., 2021), an alternative is to recommend *salad* and *spinach* which has also been consumed by the nearest neighbor $user_1$.

The idea of such a recommender could be to create diversity in terms of identifying items (or recipes) the current user did not consume up to now and—at the same time—to take into account nutritional values, i.e., to prefer items with high nutritional values (e.g., *salad* or *spinach*). Just recommending *salad* as a main dish would not be satisfactory for the user—in this situation, we can extend our basic collaborative filtering with a knowledge-based approach that supports the generation of *bundles* taking, for example, into account upper bounds in terms of the number of calories consumed per day (Beladev et al., 2016).

3.4. Quality education

Ensuring inclusive and equitable quality education and lifelong learning opportunities requires the inclusion of modern communication technologies as well as corresponding personalization concepts which help to tailor learning contents in such a way that learners can have a personalized learning experience (Klašnja-Milićević et al., 2015).

An example of applying group recommender systems in e-learning contexts on the macro level is policy decision making regarding the establishment of a new study program at a university. In such a scenario, alternative study programs could be discussed by a group of responsible stakeholders where each stakeholder can provide related proposals him/herself and can give feedback on the other existing proposals/ideas simply by evaluating the interest dimensions *feasibility* (are the personal resources available for teaching the new courses?) and *interest* (will students be interested in enrolling in the new study program?; see Table 9). We assume an evaluation scale [1..10] 1 indicating low and 10 indicating high feasibility/interest.

If we assume an equal importance of the interest dimensions feasibility and interest, the *AI* (*Artificial Intelligence*) study program could be recommended to the stakeholders since it has the highest average (AVG) evaluation. A more detailed discussion on the utility-based evaluation of alternative solutions (items, products) can be found in Felfernig et al. (2006, 2018).

On the micro-level, there exist a couple of recommendation approaches supporting the recommendation of learning items (Ribeiro, 2011; Klašnja-Milićević et al., 2015). On the one hand, content-based filtering can be applied in situations where new learning items are available for learners who are interested in a

TABLE 10 Example dataset regarding the correctness of student answers to test questions q_i (1 = correct, 0 = incorrect answer to a question q_i).

Student	$topic_1$		$topic_2$		$topic_3$	
	q_1	q_2	q_3	q_4	q_5	q_6
s_1	1	0	1	1	1	0
s_2	1	1	0	0	1	0
s_3	1	0	1	0	0	0
Correct (%)	1.0	0.33	0.66	0.33	0.66	0

TABLE 11 Example of stakeholder-specific evaluations of the qualification of different job applicants.

Stakeholder	$candidate_1$	$candidate_2$	$candidate_3$	$candidate_4$
s_1	10	5	6	7
s_2	2	7	8	8
s_3	3	7	7	6
s_4	5	8	5	7
Avg	5.0	6.75	6.5	7.0

longterm learning experience regarding a specific topic. This is similar to news recommendation where news gets recommended to users with a corresponding topic-wise reading preference. In the context of university courses, students can estimate their topic-wise expertise by answering corresponding test questions (Stettinger et al., 2020). For those topics with a lower knowledge level, content-based recommendation can be used to recommend topic-specific contents ranked on the basis of their complexity level (see Table 10).

If we assume that Table 10 is a result of a student pre-test questionnaire, the corresponding correctness shares can be used to rank the questions with regard to their complexity. For questions answered incorrectly, corresponding learning contents can be recommended, for example, by a content-based match between question category names and corresponding content categories. For example, student s_3 did not answer any question of $topic_3$ correctly. Consequently, contents related to questions q_5 and q_6 can be recommended (first, learning contents related to q_5 since this appears to be a slightly easier topic when following the *correctness* criteria).

3.5. Gender equality

The underlying goal is to *achieve gender equality and to empower all women and girls*. A major aspect in the context of achieving gender equality is the concept of fairness in terms of a gender-independent equal treatment. In recommender systems, fairness aspects play an important role in terms of assuring this property with regard to stakeholders (Li et al., 2023), for example, in music streaming platforms, musicians are interested in having their songs played and users in maximizing their positive song experience.

We expect the availability of different metrics (criteria) that help to analyze the degree to which fairness aspects have to be taken into account as well as pointing out possibilities to counteract unfair treatments (Stray et al., 2021; Wu et al., 2023). Examples thereof are *equal opportunity* requiring the same share of true positives for individual recommender system users or groups, *envy-freeness* indicating to which extent individual users or groups prefer their recommendations over the recommendations given to other users or groups, and *demographic parity* indicating that recommendations should be similar around an attribute such as *gender* (Wu et al., 2023). A simple example of how to measure the equal opportunity parity (on a scale [0..1]) of a job recommender is provided in Formula 3.

$$fairness = 1 - |accurracy(male) - accurracy(female)| \quad (3)$$

There are different ways of assuring fairness (Sonboli et al., 2022) ranging from (1) the pre-processing of a dataset on the basis of imputation, (2) the provision of fairness-aware algorithms (e.g., on the basis of integrating fairness into machine learning regularization terms), and (3) the post-processing of generated recommendations (e.g., on the basis of re-ranking recommendations). An example of assuring fairness in a group recommendation scenario (job candidate selection) is depicted in Table 11.

In the scenario shown in Table 11, stakeholders s_i are in charge of selecting a person for a specific job. In this context, a basic group recommender system is applied to recommend candidates to the group (on the basis of an *avg* aggregation function). In this example, $candidate_4$ has the best overall evaluation which could make him/her the best candidate, however, there is a strong imbalance with regard to the evaluations of $candidate_1$. For this reason, a final decision should not be taken immediately, but discussions need to be triggered regarding the contradicting evaluations of $candidate_1$. Fairness-awareness in this context means to pro-actively figure out potential issues in the decision making process in order to avoid sub-optimal decisions. An important aspect in the context of assuring fairness is also to introduce transparency into decision processes. For example, Tran et al. (2019) compare different group recommender user interfaces (differing in terms of decision process transparency) and corresponding stakeholder behaviors in terms of trying to manipulate decision outcomes. A related result is that transparency can help to counteract decision manipulation and thus to reduce the probability of sub-optimal decisions.

3.6. Clean water and sanitation

Cornerstones for the *availability of clean water and sanitation* are intelligent systems supporting the planning, implementation, and operation of corresponding technical infrastructures (Mahmoud et al., 2013; Magalhães et al., 2019).

Water management as a whole heavily relies on knowledge about the location-specific quality of water resources which is highly relevant for decision makers, involved in tasks such as land development planning. To identify relevant locations and also to predict the development of water sources over time, recommender systems can help to predict, for example, the pH level—for

TABLE 12 Simplified household water consumption data as a basis for recommending changes in consumption behavior (for shower, bathtub, toilet, and kitchen, the data describes liter p.a.).

Household	Adults	Children	Shower	Bathtub	Toilet	Kitchen
h_1	2	2	1,000	20,000	2,000	800
h_2	2	0	300	12,000	1,200	1,000
h_3	2	2	4,000	30,000	2,500	1,000

related details on an example application we refer to Mahmoud et al. (2013). Related techniques for designing relevant sanitation concepts are also in the need of a decision support able to integrate local decision makers (Magalhães et al., 2019).

In the context of optimizing household water consumption, recommender systems can be applied to sensitize users in terms of adapting, i.e., reducing their water consumption (Arsene et al., 2023). Table 12 provides a simple example dataset representing different households with corresponding consumption data. Our assumption in this context is the availability of smart-meter technologies allowing the measurement of water consumptions with individual water devices.

In this example (Table 12), despite an equivalent number of persons living in the household, household h_3 has a significantly higher water consumption compared to household h_1. Household h_1 can be regarded as a nearest neighbor of household h_3. The corresponding differences in consumption can be used as a basis for generating corresponding explanations (Arsene et al., 2023). Depending on the water device specific differences, recommendations can propose actions such as taking shorter showers, using lower-flow shower-heads, and turning off taps during tooth-brushing (Arsene et al., 2023).

3.7. Affordable and clean energy

The major related goal is the *provision of affordable, reliable, sustainable, and modern energy for all*. Recommender systems can help in the establishment of related energy provision infrastructures such as wind energy systems with layout planning (Sultana et al., 2022) and related performance optimizations (Vaghasiya et al., 2017; Pinciroli et al., 2022). Achieving the goal of supporting affordable and clean energy also requires the support of public campaigns that indicate in the form of explanations and argumentations which behavior patterns can help to reduce individual energy consumption which is a major goal of assuring affordable and clean energy (Starke et al., 2021). A similar scenario has already been discussed within the scope of the goal of *good health and wellbeing*, i.e., a recommender system can be applied to personalize related messages. Message personalization requires the availability of basic user data such as type of home (e.g., apartment vs. own house), number of family members, and further information regarding personal energy consumption patterns (Eirinaki et al., 2022) and also knowledge about persuasive technologies (Adaji and Adisa, 2022) and effective user interfaces (Starke et al., 2017) to achieve sustainable behavior.

On the level of individual households, energy efficiency can be achieved on the basis of household-specific energy breakdowns (Batra et al., 2017; Himeur et al., 2021). In this context, recommendation techniques of collaborative filtering and matrix factorization can help to predict the energy consumption of households who did not perform a breakdown up to now, for example, for reasons of related costs (Batra et al., 2017). Household-specific energy consumption can also be triggered on the basis of comparative and community-based explanations (Petkov et al., 2011) where the energy saving performance of individual households can be compared to each other indicating personal performances compared to other households. *Norm-based comparisons* are an example thereof: *the majority of similar households show a better energy saving compared to your current savings data*. Furthermore, explanations can refer to energy consumption in the past (*self-comparison* feedback) and indicate improvement or deterioration.

3.8. Decent work and economic growth

The underlying goal is to *promote economic growth, full and productive employment, and decent work for all*. Nowadays, recommender systems can be regarded as a core technology helping to further increase the business value of offered products and services (Jannach and Jugovac, 2019). Examples of related measurements are *click-through rates* and *sales/revenue*. However, recommender systems supporting sustainability development goals have a different focus. For example, the impact of recommender systems on increasing the quality of education can be measured directly in terms of increased knowledge levels of different social groups. Furthermore, the impact of recommender systems in the context of clean energy and energy savings can be measured, for example, in terms of reduced household-wise energy consumption. Consequently, for achieving sustainability goals, evaluation metrics should be more customer-focused and thus also *consequence-based* compared to metrics in standard business scenarios.

Recommender systems can also help to improve the quality of work and sustainable growth in terms of supporting different kinds of open innovation processes. Achieving sustainability goals is a central agenda of public administrations and finding relevant acceptable solutions for achieving these goals has to be performed in terms of a participatory innovation and design process (Felfernig et al., 2004; Brocco and Groh, 2009; Smith and Iversen, 2018; Shadowen et al., 2020). In this context, recommender systems can be applied to support idea generation processes, for example, by recommending ideas to community members interested in similar topics (Haiba et al., 2017).

Recommender systems are an established technology in different people to people (P2P) recommendation scenarios—examples thereof are recommending new friends in social networks, recommending business partnerships, and recommending jobs (Gutiérrez et al., 2019; Koprinska and Yacef, 2022). Finding the right job is crucial for a further personal development and a productive employment. In these scenarios, recommender systems support a matchmaking functionality by "connecting" job offers with interested employees. Often, such scenarios are based on content-based recommendation where job descriptions are matched with the interest and qualification profiles of potential candidates. An important issue in these scenarios is the aspect of fairness with regard to both, institutions offering a job and corresponding candidates. From the institution point of view, fairness should be guaranteed with respect to other institutions offering similar jobs, i.e., amount and expertise of contacted candidates should be nearly the same. From the candidates point of view, no overloading should take place, i.e., a specific job offer should not be shared with all potential candidates. Finally, a stable or increasing number of new established enterprizes can be regarded as a major indicator of economic growth (Luef et al., 2020)—in this context, recommender systems can be applied to support investors in better identifying the most relevant investments.

3.9. Industry, innovation, and infrastructure

The underlying goal is to promote *innovation, sustainable industrialization, and resilient infrastructures*. Industrial applications of recommender systems are many-fold and range from the recommendation of movies (Gomez-Uribe and Hunt, 2016), the recommendation of books (Smith and Linden, 2017), recommendations in the dating business (Tomita et al., 2022), to the recommendation of airline offers (Dadoun et al., 2021). Beyond acting as a support of core business processes (e.g., selling books), recommender systems can also act in a supportive role which is often the case with sustainability topics.

Recommender systems can be applied as a knowledge transfer medium for different industrial segments to indicate possibilities in terms of process improvements and the inclusion of sustainable materials into production processes (Wiezorek and Christensen, 2021). Identifying sustainability properties of products is often not an easy task—examples of such properties are environmental impact, animal welfare, and customer benefits (Tomkins et al., 2018). Due to a lack of easily accessible sustainability information, customers do not always behave as intended, i.e., although interested in sustainability, they take sub-optimal decisions due to the lack of related information. Tomkins et al. (2018) introduce a hybrid recommender system where the item-related sustainability classification is based on probabilistic soft logic.

Fostering innovation can be supported in various forms—examples thereof are *innovation processes* where recommender systems provide support in the configuration of innovation teams, i.e., who should work together to achieve specific innovation goals (Brocco and Groh, 2009) and the *process of idea generation* (Haiba et al., 2017). An important aspect in software development is to overcome the barriers of taking into account sustainability aspects in software engineering (Roher and Richardson, 2013). Also in this context, recommender systems can be applied to support project stakeholders with recommendations that are determined depending on the underlying application domain. Similar applications exist in software development, where intelligent source code analysis can help to identify software elements to be adapted, for example, to achieve more efficient runtimes and corresponding CPU usage (Muralidhar et al., 2022).

3.10. Reduced inequalities

Achieving this objective (*reduce inequality within and among countries*) requires actions such as promoting economic inclusion, direct investments, and fostering mobility and migration to bridge divides.

On the macro-level, recommender systems can help to figure out new potentials overlooked by countries, that can trigger future economic welfare due to strategic future advantages (Liao et al., 2018). In this line of research, recommender systems can also help to establish new study programs of relevance helping to promote relevant know-how for implementing specific industries. As discussed in Che (2020), recommender systems can be applied in the context of developing export diversification strategies resulting in recommended industry/product segments which should be expanded or established in specific countries. Having identified such segments, recommender systems can also be applied to identify a corresponding educational focus indicating which study programs should be emphasized or established in a specific country or a specific region (Tavakoli et al., 2022).

Specifically in the context of fostering mobility and migration, the task of country recommendation becomes increasingly relevant. Majjodi et al. (2020) motivate the application of country recommender systems since beginning a new life in a different country is for various reasons a high-involvement and often risky decision. The basic underlying idea is to support country recommendation on the basis of collaborative filtering where preferences of existing emigrants are used to infer relevant countries for potential emigrants. Such a scenario can typically not be supported solely on the basis of collaborative filtering (which relies on medium- and long-term preferences) but must include a knowledge-based recommendation component that takes into account short-term circumstances, for example, changing political situations, which do not allow a corresponding recommendation. This is a typical example of hybrid recommendation, where synergy effects of different recommenders can be combined in a reasonable fashion (Burke, 2002).

Fairness aspects play a crucial role in different job recommendation scenarios (Li et al., 2023). In such scenarios, job candidates should receive recommendations with a very good fit but at the same time companies offering jobs should be treated equally in terms of amount and quality of proposed candidates. A related simplified recommendation scenario is depicted in Table 13. Table 13 shows individual job candidate/job compatibilities determined, for example, on the basis of content-based recommendation which provides a similarity between a job

TABLE 13 Simplified example of taking into account fairness aspects in job recommendation scenarios.

Candidate	job_1	job_2	job_3	job_4	job_5	job_6
c_1	9	9	8	1	8	1
c_2	9	1	7	9	2	7
c_3	2	1	6	8	7	2

TABLE 14 Recommendations of candidate/job assignments where 1 (in brackets) indicates that the corresponding assignment is part of the recommendation *REC*.

Candidate	job_1	job_2	job_3	job_4	job_5	job_6
c_1	9 (1)	9 (1)	8 (1)	1 (0)	8 (1)	1 (0)
c_2	9 (1)	1 (0)	7 (0)	9 (1)	2 (0)	7 (1)
c_3	2 (0)	1 (0)	6 (0)	8 (1)	7 (0)	2 (0)

description and the application material provided by the candidate (in our example, on a scale [1..10]—the higher the better).

In this setting, different fairness aspects can be taken into account. For example, each candidate should have at least one job offering (see Formula 4).

$$\forall c \in candidates : \#jobs(c) > 0 \tag{4}$$

Furthermore, there should be at least one candidate for each job offering (see Formula 5).

$$\forall j \in jobs : \#candidates(j) > 0 \tag{5}$$

Finally, the recommendation quality should be maximized where *REC* denotes the set of all proposed job/candidate assignments ($rec \in REC$) and *maxrating* is the maximum (best) possible candidate/job rating. In this context, the optimization goal is to *minimize* the average distance between candidate/job compatibility evaluations and the maximum possible rating (see Formula 6).

$$MIN \leftarrow \frac{\Sigma_{rec \in REC}\ maxrating - rating(rec)}{|REC|} \tag{6}$$

A recommendation *REC* for candidate/job assignments on the basis of the example scenario shown in Table 13 is presented in Table 14.

In this example, *REC* consists of 8 proposed assignments where candidate c_1 is recommended for four jobs ($job_1, job_2, job_3, job_5$), c_2 is recommended for three jobs (job_1, job_4, job_6), and c_1 for one job (job_4).

Finally, fairness considerations are also relevant in the context of individuals with disabilities. Related recommendation approaches support content recommendation (Quisi-Peralta et al., 2018; Apostolidis et al., 2022), recommendation for accessibility and mobility (Cardoso et al., 2015; Brodeala, 2020; Tsai et al., 2022), activity recommendation (Altulyan et al., 2019), and the recommendation of points of interest (Mauro et al., 2022).

3.11. Sustainable cities and communities

The related goal is to *make cities and human settlements inclusive, safe, resilient, and sustainable*. City planners, decision makers, and citizens need to be supported in order to achieve the different goals of sustainable cities and communities. For example, sustainable mobility provides modern commuting systems and facilities on the basis of green infrastructures. Furthermore, in order to assure a smart environment, natural resources need to be preserved.

Recommender systems can support sustainable smart cities on the basis of supporting strategic decision making. Depending on the context of a specific city, different actions need to be taken in order to be able to achieve related sustainable development goals (Bokolo, 2021). Helping public stakeholders to achieve related sustainability goals can be supported, for example, on the basis of case-based recommender systems which follow the idea of supporting the identification of similar cases (cities) and on the basis of related measures already completed in similar cities to recommend sustainability-fostering activities for the current city (Banerjee, 2023).

In such contexts, recommender systems can support also individuals (e.g., citizens and tourists) in the completion of their tasks and the achievement of their goals. For example, sustainable tourism recommender systems are able to propose relevant points of interest (POI) whilst taking into account aspects such as negative environmental impacts, local communities, and cultural heritage (Khan et al., 2021; Banerjee, 2023; Merinov, 2023). Related interventions are needed that assure fairness among multiple stakeholders such as tourists, tourism organizations, local citizens, and environmental aspects such as water quality, air quality, and wildlife. Calculating recommendations in such scenarios requires the integration of optimization methods supporting, for example, the optimization of round trips of individual travel groups, resource balancing in the sense that not too many tourists visit specific sightseeing destinations at the same time (triggering issues in terms of disturbances, environmental pollution, and the scaring of animals; Sihotang et al., 2021; Merinov, 2023). In such contexts, explanations can help to assure recommendation understandability and to sensitize stakeholders with regard to sustainability aspects (Banerjee, 2023).

3.12. Responsible consumption and production

The underlying goal is to *ensure sustainable consumption and production patterns*. A challenge in this context is to find ways to achieve environment sustainability and at the same time to trigger economic growth and welfare by making these two factors much more independent, i.e., to "achieve more with less."

Sustainable production is related to the goal of achieving industrial symbioses where cooperations between companies are intensified, for example, with the goal to minimize industrial waste streams and share related knowledge (van Capelleveen et al., 2018). In such contexts, recommender systems can support individual companies by the recommendation of opportunities in waste marketplaces which in the following could lead to intensified cooperations between companies. In such scenarios, recommender systems must be built in a knowledge-based fashion which helps to assure that the needed knowledge about compatibilities of waste products is available. Such basic recommendations can be enhanced by future recommender systems proposing different types of cooperations based on deep knowledge about the underlying waste chains. We regard this scenario as part of the macro level (in the case that public agencies deliver related recommendations for companies) and on the micro-level, if companies themselves are registered in a public marketplace.

Achieving sustainability goals in the fashion industry (Wu et al., 2022) requires, for example, to lower the number of returned deliveries and to increase a customers willingness to accept higher prices for higher-quality items. Such goals can be achieved, for example, by providing means to create bundles of items (Li et al., 2020; Wiezorek and Christensen, 2021) which fit together relieving customers from the burden of performing this task on their own (Zielnicki, 2019). In this context, persuasive explanations are needed that help to better motivate customers to choose more sustainable options (Knowles et al., 2014). An important aspect is also to assure solution minimality, i.e., to guarantee that product bundles and complex configurations do not entail unnecessary components (Vidal-Silva et al., 2021).

3.13. Climate action

The major related challenge is to perform actions with the goal to *combat climate change and direct or indirect impacts thereof*. An important aspect in combating climate change is to empower new types of energy production systems, for example, in terms of prosumer networks where private households can act as solar energy producers and consumers at the same time (Guzzi and Chiodo, 2022). Before establishing individual cooperations, it is important to figure out and recommend homogeneous prosumer clusters which then maximize the consumption of the cluster-produced energy and—at the same time—minimize the consumption of external energy sources. Recommendations in this context can propose specific clusters in a region of consumers (Guzzi and Chiodo, 2022). In related energy saving scenarios, persuasive explanations of recommendations play a central role since households should be encouraged to reduce energy consumption in a sustainable fashion. Starke et al. (2021) show how such explanations can be designed on the basis of the concepts of *framing* (Tversky and Kahneman, 1985) where those attributes of a decision alternative are highlighted in a recommender user interface which are related to high kWh savings. One simple possibility of "implementing" framing on the user interface level is to sort recommended items on specific attributes making those items more attractive that score high with regard to this attribute. For example, alternative energy saving measures can be sorted with regard to the amount of kWh savings (Starke et al., 2021). These insights regarding the provision of explanations can also be applied in public services provision when informing citizens about potential energy saving measures. Besides the mentioned energy saving scenarios, such persuasive messaging can also be applied in the context of route recommendation scenarios with the goal to encourage users to choose environmental-friendly routes thus contributing to reduce pollution due to carbon emissions (Bothos et al., 2016).

On the level of individual households, recommender systems can be applied to assist residents in optimizing energy savings. Supporting such optimizations, is a central capability of constraint-based recommender systems (Felfernig and Burke, 2008) which allow the inclusion of optimization criteria to determine relevant recommendation candidates (Murphy et al., 2015). If, for example, power suppliers, support time-dependent flexible pricing conditions, the operation of electric equipment should be optimized on the basis of the pricing models. Furthermore, such constraint-based applications can take into account corresponding regional weather forecasts and conditions to also take into account potential consumptions of energy produced by the household itself thus supporting real-time recommendations and corresponding actions in terms of activating and deactivating a specific heating equipment (Dahihande et al., 2020). An important aspect is also that the recommender has knowledge about the current in-building location of residents. Using such knowledge, can help to further decrease power consumption in buildings by activating/deactivating electronic equipment in an intelligent fashion (Wei et al., 2018).

3.14. Life below water

The underlying goal is to *enable a sustainable use of oceans, seas, and marine resources*. The application of artificial intelligence techniques in related fields is progressing, however, there is potential for further machine learning and recommender systems applications (Xu et al., 2022).

Water quality and pollution assessment and the development of countermeasures becomes an increasingly relevant issue. Due to limited resources in terms of possible data collections and available datasets, machine learning models need to be developed that serve as a basis for pollution prediction but also the determination of recommendations of relevant counter-measures (Xu et al., 2022). In the context of illegal fishing, recommender systems can help to propose effective sequential defender strategies that help to counteract illegal fishing (Fang et al., 2015).

A relevant problem directly related to water quality and further environmental conditions is the provision of recommendations for aquacultures (e.g., fish farming), for example, in terms of species suitable for the specific conditions and also in terms of nutrients that should be provided in such contexts (Praba et al., 2023). Related recommender applications can also be applied for further tasks, for example, identification and counteracting fish diseases, remote maintenance of offshore infrastructures, and recommending nutrition plans depending a.o. on estimated weight and size of fishes.

3.15. Life on land

The overall underlying goal is *a sustainable use of terrestrial ecosystems*, for example, in terms of *sustainability in forest* *management, counteracting desertification, and halting of biodiversity loss.*

It is important to understand and optimally decide on appropriate crops to be cultivated. Crop recommender systems recommend crops on the basis of land quality and mineral requirements whereas pesticide recommender systems propose a collection of pesticides in order to protect specific crops from diseases (Patel and Patel, 2020; Usman et al., 2021). In the line of sustainability requirements, such systems have to take into account impacts of potentially used treatments (e.g., pesticides), i.e., not solely focusing on maximizing productivity but trying to keep soil characteristics are extremely important for maintaining fertility (Usman et al., 2021). In a broader sense, recommender systems can be applied to support different kinds of precision farming (Ronzhin et al., 2022; Thilakarathne et al., 2022; Wakchaure et al., 2023).

Furthermore, recommender systems can provide suggestions on how to counteract wildlife poaching which is a serious extinction threat to many animal species and related ecosystems (Nguyen et al., 2016). Based on such tools, animal protectors are enabled to analyze and predict poaching activities and to recommend countermeasures on the basis of behavioral models learning from poaching data (Yang et al., 2014; Nguyen et al., 2016). In this context, resource balancing plays an important role since personal resources used for observation activities are extremely limited (Yang et al., 2014).

3.16. Peace, justice, and strong institutions

The underlying goal is to *promote peaceful societies supporting justice for all on the basis of corresponding effective, accountable, and inclusive institutions*. Law enforcement agencies are aware of the fact that the analysis of networks of co-offenders who committed crimes together is highly relevant in crime investigation (Tayebi et al., 2011). Manually performing such tasks can be quite inefficient which make it an application scenario for recommender systems: suspects are compared with known co-offending networks and the most relevant ones are shown (recommended) to the law enforcement agency representatives.

In the context of trials, recommender systems can support legal practitioners in the identification of advantageous arguments for an ongoing case (Dhanani et al., 2021). In practice, documents and further material related to the current case are compared with already "closed" cases on the basis of different text-based similarity metrics. The identified most similar documents are then used as a basis for more detailed analysis steps conducted with the goal of identifying relevant arguments better helping to win acquittal for an accused person (Mandal et al., 2017; Dhanani et al., 2021). On the negative side, such content-based recommenders are also applied by different social media and news platforms with the danger of creating so-called "echo-chambers" of misinformation (Sallami et al., 2023)—this is also related to the general requirement of considering and minimizing harm in recommenders (Ekstrand and Ekstrand, 2016).

3.17. Partnerships for the goals

The goal is to *identify global partnerships bringing together various institutions such as governments, private sector, and others that help to better achieve the discussed goals*. A specific task is to assure an increasing support for developing countries to assure an equitable progress for all and also strengthen the path toward sustainability. Identifying and establishing such cooperations can also be supported by recommender systems, for example, people-2-people recommender systems can support the identification of business partners and research cooperations (Hu and Ma, 2021; Koprinska and Yacef, 2022).

4. Open research issues

4.1. Evaluation metrics for sustainability

There exists a plethora of evaluation metrics for recommender systems (Zangerle and Bauer, 2022) ranging from (1) data-driven approaches to evaluate the prediction/classification quality, (2) experimental settings evaluating prototype systems with alternative variants of user interfaces and algorithmic approaches, and (3) field studies in real-world settings, for example, on the basis of A/B testing. However, existing evaluation metrics do not focus on specific sustainability aspects, for example, achievements in terms of reduced power consumption, increased share of sustainable items in a user's purchase history, and reduced global CO_2 footprint—a specific related aspect is to take sustainability aspects into account when selecting and/or implementing recommendation algorithms (Lannelongue et al., 2023; Spillo et al., 2023).

4.2. Nudging for sustainability

The way decision alternatives are presented to users has an impact on the final decisions taken by users. In this context, *nudging* (Thaler and Sunstein, 2021) can be defined as any aspect of a choice situation that alters the behavior of a user in a predicable way without forbidding any options. Providing a basis for better choice on the basis of decision support is an important goal to be taken into account (Kroese et al., 2015). Related research already indicates the potential of nudges in various recommender systems supporting sustainability goals (Bothos et al., 2015; Lehner et al., 2016; Karlsen and Andersen, 2019; Majjodi et al., 2022). Successful nudges are often based on decision biases, i.e., decision practices (heuristics) used by humans to often lead to suboptimal decision outcomes. An overview of such decision biases and their role in recommender systems is discussed in Mandl et al. (2011), Chen et al. (2013), Lex et al. (2021), and Tran et al. (2021).

4.3. Contextual explanations

Given an infrastructure of intelligent data collection, energy consumption information is directly available and can be used for generating corresponding recommendations. For example, in smart homes the activation of a dishwasher and a washing machine could be delayed due to the fact that a parallel car battery recharging would lead to an additional consumption of external energy resources. In travel scenarios, a recommender system can detect alternative (more sustainable) routes not requiring a car rental. In such scenarios, explanations play an important role and must be contextualized and personalized to attain the maximum impact. Explanation generation for achieving sustainability goals can be regarded as a highly relevant research issue (Starke et al., 2021).

4.4. Consequence-based explanations

In the context of recommender systems, explanations can be used to support different goals such as trust and persuasiveness (in terms of increasing the probability that a user will purchase an item; Tintarev and Masthoff, 2012). However, with a few exceptions, existing explanation approaches do not take into account the consequences of "accepting" a recommendation. For example, purchasing a rather expensive *BMW* has specific consequences on the economic situation of a household—having an expensive car could have an impact on the affordability of holidays or the education quality of children. Specifically in the context of achieving sustainability goals, there is a need to analyze alternatives in terms of the corresponding consequences. For example, explanations can provide information regarding the consequences of not investing into new heating equipment [in terms of CO_2 footprint issues as well as in terms of additional costs associated with the old (still installed) heating equipment].

4.5. Constraint-based recommendation for sustainability

Constraint-based approaches are applied in various contexts, for example, the optimization of a households energy consumption strategy (Murphy et al., 2015). In the line of the idea of simulating the consequences of financial decisions (Fano and Kurth, 2003), constraint-based recommenders could also be combined with corresponding simulation components that help to visualize the impact of different decisions. For example, sticking with the old heating equipment could have an impact on the overall related costs in the long run. Furthermore, consequences exist on different levels, for example, related simulations could also represent "what-if" scenarios, i.e., what happens to the global warming if a majority of people are not thinking about reducing their CO_2 footprint.

5. Conclusions

Sustainability development goals (SDGs) as defined by the United Nations are a call for action to planet protection, ending poverty, and ensuring peace and prosperity. In this article, we have provided an overview of SDGs and related applications of recommender systems. These systems can be regarded as a core technology of different decision support scenarios and thus play a major role in achieving the mentioned SDGs. In order to assure understandability, we have provided corresponding working examples that show how recommender systems can be applied in different application contexts. Furthermore, with the goal to foster further related research, we have provided a list of research issues in the context of developing recommender systems supporting sustainability goals.

Author contributions

AF: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing. MW: Conceptualization, Methodology, Project administration, Resources, Writing—original draft, Writing—review and editing. TT: Conceptualization, Writing—original draft, Writing—review and editing. SP-E: Conceptualization, Investigation, Writing—original draft, Writing—review and editing. SL: Conceptualization, Writing—original draft, Writing—review and editing. ME: Conceptualization, Writing—review and editing. DG: Conceptualization, Writing—review and editing. V-ML: Conceptualization, Writing—review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The presented work has been developed within the TU Graz internal project AI4SUSTAINABILITY. This work was supported by TU Graz Open Access Publishing Fund.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer TD declared a shared affiliation with the author(s) to the handling editor at the time of review.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abu-Issa, A., Hajjaj, S., Al-Jamal, S., Barghotti, D., Awad, A., Hussein, M., et al. (2023). "Design and implementation of proactive multi-type context-aware recommender system for patients suffering diabetes," in *International Conference on Smart Applications, Communications and Networking (SmartNets)* (New York, NY), 1–7. doi: 10.1109/SmartNets58706.2023.10216111

Adaji, I., and Adisa, M. (2022). "A review of the use of persuasive technologies to influence sustainable behavior," in *30th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '22 Adjunct* (New York, NY), 317–325. doi: 10.1145/3511047.3537653

Alcaraz-Herrera, H., Cartlidge, J., Toumpakari, Z., Western, M., and Palomares, I. (2022). EvoRecSys: evolutionary framework for health and well-being recommender systems. *User Model. User-Adapt. Interact.* 32, 883–921. doi: 10.1007/s11257-021-09318-3

Altulyan, M., Huang, C., Yao, L., Wang, X., Kanhere, S., and Cao, Y. (2019). "Reminder care system: an activity-aware cross-device recommendation system," in *Advanced Data Mining and Applications*, eds J. Li, S. Wang, S. Qin, X. Li, and S. Wang (Cham: Springer), 207–220. doi: 10.1007/978-3-030-35231-8_15

Apostolidis, K., Mezaris, V., Papadogiorgaki, M., Bei, E., Livanos, G., and Zervakis, M. (2022). Content and other resources recommendations for individuals with intellectual disability: a review. *Electronics* 11:3472. doi: 10.3390/electronics11213472

Arévalo, P., Orellana, M., Cedillo, P., Lima, J.-F., and Zambrano-Martinez, J. L. (2022). "A methodology to develop an outdoor activities recommender based on air pollution variables," in *Information and Communication Technologies*, eds J. Herrera-Tapia, G. Rodriguez-Morales, C. E. R. Fonseca, and S. Berrezueta-Guzman (Cham: Springer), 171–185. doi: 10.1007/978-3-031-18272-3_12

Arsene, D., Predescu, A., Truică, C., Apostol, E., and Mocanu, M. (2023). Decision support strategies for household water consumption behaviors based on advanced recommender systems. *Water* 15, 1–13. doi: 10.3390/w15142550

Balassa, B., and Noland, M. (1989). "Revealed" comparative advantage in Japan and the United States. *J. Int. Econ. Integr.* 4, 8–22. doi: 10.11130/jei.1989.4.2.8

Banerjee, A. (2023). "Fairness and sustainability in multistakeholder tourism recommender systems," in *31st ACM Conference on User Modeling, Adaptation and Personalization, UMAP '23* (New York, NY), 274–279. doi: 10.1145/3565472.3595607

Batra, N., Wang, H., Singh, A., and Whitehouse, K. (2017). "Matrix factorisation for scalable energy breakdown," in *31st AAAI Conference on Artificial Intelligence, AAAI'17* (Washington, DC: AAAI Press), 4467–4473. doi: 10.1609/aaai.v31i1.11179

Beladev, M., Rokach, L., and Shapira, B. (2016). Recommender systems for product bundling. *Knowl. Based Syst.* 111, 193–206. doi: 10.1016/j.knosys.2016.08.013

Blasioli, E., Mansouri, B., Tamvada, S., and Hassini, E. (2023). Vaccine allocation and distribution: a review with a focus on quantitative methodologies and application to equity, hesitancy, and COVID-19 pandemic. *Oper. Res. Forum* 4:27. doi: 10.1007/s43069-023-00194-8

Bokolo, A. (2021). A case-based reasoning recommender system for sustainable smart city development. *AI Soc.* 36, 159–183. doi: 10.1007/s00146-020-00984-2

Bothos, E., Apostolou, D., and Mentzas, G. (2016). "A recommender for persuasive messages in route planning applications," in *7th International Conference on Information, Intelligence, Systems & Applications (IISA)* (New York, NY), 1–5. doi: 10.1109/IISA.2016.7785399

Bothos, E., Apostolou, D., Mentzas, G., Tsihrintzis, G., and Virvou, M. (2015). Recommender systems for nudging commuters towards eco-friendly decisions. *Intell. Decis. Technol.* 9, 295–306. doi: 10.3233/IDT-140223

Bouni, M., Hssina, B., Douzi, K., and Douzi, S. (2022). Towards an efficient recommender systems in smart agriculture: a deep reinforcement learning approach. *Proc. Comput. Sci.* 203, 825–830. doi: 10.1016/j.procs.2022.07.124

Brocco, M., and Groh, G. (2009). "Team recommendation in open innovation networks," in *3rd ACM Conference on Recommender Systems, RecSys '09* (New York, NY), 365–368. doi: 10.1145/1639714.1639789

Brodeala, L. (2020). "Online recommender system for accessible tourism destinations," in *14th ACM Conference on Recommender Systems, RecSys '20* (New York, NY), 787–791. doi: 10.1145/3383313.3411450

Bui, T. (2000). "Decision support systems for sustainable development," in *Decision Support Systems for Sustainable Development: A Resource Book of Methods and Applications*, eds G. Kersten, Z. Mikolajuk, and A. Yeh (New York, NY: Springer), 1–10.

Burke, R. (2000). Knowledge-based recommender systems. *Encyclop. Lib. Inform. Syst.* 69, 180–200.

Burke, R. (2002). Hybrid recommender systems: survey and experiments. *User Model. User-Adapt. Interact.* 12, 331–370. doi: 10.1023/A:1021240730564

Cappella, J., Yang, S., and Lee, S. (2015). Constructing recommendation systems for effective health messages using content, collaborative, and hybrid algorithms. *Ann. Am. Acad. Polit. Soc. Sci.* 659, 290–306. doi: 10.1177/0002716215570573

Cardoso, I. I. G., Mota, B., Barbosa, J., and da Rosa Righi, R. (2015). "Vulcanus: a recommender system for accessibility based on trails," in *Latin American Computing Conference (CLEI)* (New York, NY), 1–12. doi: 10.1109/CLEI.2015.7360003

Che, N. (2020). "Intelligent export diversification: an export recommendation system with machine learning," in *DecisionSciRN: Intelligent Decision Support Systems (Topic)*. doi: 10.2139/ssrn.3721200

Chen, L., de Gemmis, M., Felfernig, A., Lops, P., Ricci, F., and Semeraro, G. (2013). Human decision making and recommender systems. *ACM Trans. Interact. Intell. Syst.* 3, 1–7. doi: 10.1145/2533670.2533675

Chen, L., and Pu, P. (2012). Critiquing-based recommenders: survey and emerging trends. *User Model. User Adapt. Interact.* 22, 125–150. doi: 10.1007/s11257-011-9108-6

Dadoun, A., Defoin-Platel, M., Fiig, T., Landra, C., and Troncy, R. (2021). How recommender systems can transform airline offer construction and retailing. *J. Reven. Pricing Manage.* 20, 301–315. doi: 10.1057/s41272-021-00313-2

Dahihande, J., Jaiswal, A., Pagar, A., Thakare, A., Eirinaki, M., and Varlamis, I. (2020). "Reducing energy waste in households through real-time recommendations," in *14th ACM Conference on Recommender Systems, RecSys '20* (New York, NY), 545–550. doi: 10.1145/3383313.3412212

Dhanani, J., Mehta, R., Rana, D., Thampi, S., El-Alfy, E., and Trajkovic, L. (2021). Legal document recommendation system: a cluster based pairwise similarity computation. *J. Intell. Fuzzy Syst.* 41, 5497–5509. doi: 10.3233/JIFS-189871

Eirinaki, M., Varlamis, I., Dahihande, J., Jaiswal, A., Pagar, A., and Thakare, A. (2022). Real-time recommendations for energy-efficient appliance usage in households. *Front. Big Data* 5:972206. doi: 10.3389/fdata.2022.972206

Ekstrand, J. D., and Ekstrand, M. D. (2016). "First do no harm: considering and minimizing harm in recommender systems designed for engendering health," in *Workshop on Recommender Systems for Health at RecSys '16*, 1–2.

Ekstrand, M., Riedl, J., and Konstan, J. (2011). Collaborative filtering recommender systems. *Found. Trends Hum. Comput. Interact.* 4, 81–173. doi: 10.1561/1100000009

Falkner, A., Felfernig, A., and Haag, A. (2011). Recommendation technologies for configurable products. *AI Mag.* 32, 99–108. doi: 10.1609/aimag.v32i3.2369

Fang, F., Stone, P., and Tambe, M. (2015). "When security games go green: designing defender strategies to prevent poaching and illegal fishing," in *IJCAI'15* (New York, NY: AAAI Press), 2589–2595.

Fano, A., and Kurth, S. (2003). "Personal choice point: helping users visualize what it means to buy a BMW," in *8th International Conference on Intelligent User Interfaces (IUI'03)* (New York, NY), 46–52. doi: 10.1145/604045.604057

Felfernig, A., Boratto, L., Stettinger, M., and Tkalcic, M. (2018). *Group Recommender Systems: An Introduction, 1st Edn*. Cham: Springer. doi: 10.1007/978-3-319-75067-5

Felfernig, A., and Burke, R. (2008). "Constraint-based recommender systems: technologies and research issues," in *Proceedings of the 10th International Conference on Electronic Commerce, ICEC '08* (New York, NY), 1–10. doi: 10.1145/1409540.1409544

Felfernig, A., Friedrich, G., Jannach, D., and Zanker, M. (2006). An integrated environment for the development of knowledge-based recommender applications. *Int. J. Electron. Commerce* 11, 11–34. doi: 10.2753/JEC1086-4415110201

Felfernig, A., Russ, C., and Wundara, M. (2004). "Toolkits supporting open innovation in E-Government," in *6th International Conference on Enterprise Information Systems (ICEIS2004), Vol. 4*, eds I. Seruca, U. Filipe, S. Hammoudi, and J. Cordeiro (INSTICC Press), 296–302.

Gill, H., Sehgal, V., and Verma, A. (2021). "A context aware recommender system for predicting crop factors using LSTM," in *2021 Asian Conference on Innovation in Technology (ASIANCON)* (New York, NY), 1–4. doi: 10.1109/ASIANCON51346.2021.9544692

Gomez-Uribe, C., and Hunt, N. (2016). The netflix recommender system: algorithms, business value, and innovation. *ACM Trans. Manage. Inform. Syst.* 6:4 doi: 10.1145/2843948

Gutiérrez, F., Charleer, S., Croon, R. D., Htun, N., G, G., and Verbert, K. (2019). "Explaining and exploring job recommendations: a user-driven approach for interacting with knowledge-based job recommender systems," in *13th ACM Conference on Recommender Systems, RecSys '19* (New York, NY), 60–68. doi: 10.1145/3298689.3347001

Guzzi, P., and Chiodo, F. (2022). Towards a recommender system for profiling users in a renewable energetic community. *ArXiv:abs/2209.05465*. Ithaca, NY.

Haiba, M. E., Elbassiti, L., and Ajhoun, R. (2017). Using recommender systems to support idea generation stage. *J. Eng. Appl. Sci.* 12, 9341–9351. doi: 10.36478/jeasci.2017.9341.9351

Herpich, M., Rist, T., Seiderer, A., and André, E. (2017). "Towards a gamified recommender system for the elderly," in *International Conference on Digital Health, DH '17* (New York, NY), 211–215. doi: 10.1145/3079452.3079500

Himeur, Y., Alsalemi, A., Al-Kababji, A., Bensaali, F., Amira, A., Sardianos, C., et al. (2021). A survey of recommender systems for energy efficiency in buildings: principles, challenges and prospects. *Inform. Fus.* 72, 1–21. doi: 10.1016/j.inffus.2021.02.002

Hu, D., and Ma, H. (2021). Collaborator recommendation integrating author's cooperation strength and research interests on attributed graph. *Adv. Comput. Intell.* 1:2. doi: 10.1007/s43674-021-00002-y

Jannach, D., and Jugovac, M. (2019). Measuring the business value of recommender systems. *ACM Trans. Manage. Inf. Syst.* 10:4. doi: 10.1145/3370082

Julia, C., Fialon, M., Galan, P., Deschasaux-Tanguy, M., Andreeva, V., et al. (2021). Are foods "healthy" or "healthier"? Front-of-pack labelling and the concept of healthiness applied to foods. *Br. J. Nutr.* 127, 948–952. doi: 10.1017/S0007114521001458

Karlsen, R., and Andersen, A. (2019). Recommendations with a nudge. *Technologies* 7:45. doi: 10.3390/technologies7020045

Khan, R., Lim, C., Ahmed, M., Tan, K., and Mokhtar, M. (2021). Systematic review of contextual suggestion and recommendation systems for sustainable e-tourism. *Sustainability* 13:8141. doi: 10.3390/su13158141

Klašnja-Milićević, A., Ivanović, M., and Nanopoulos, A. (2015). Recommender systems in e-learning environments: a survey of the state-of-the-art and possible extensions. *Artif. Intell. Rev.* 44, 571–604. doi: 10.1007/s10462-015-9440-z

Knowles, B., Blair, L., Walker, S., Coulton, P., Thomas, L., and Mullagh, L. (2014). "Patterns of persuasion for sustainability," in *2014 Conference on Designing Interactive Systems, DIS '14* (New York, NY), 1035–1044. doi: 10.1145/2598510.2598536

Koprinska, I., and Yacef, K. (2022). "People-to-people reciprocal recommenders," in *Recommender Systems Handbook*, eds F. Ricci, L. Rokach, and B. Shapira (Cham: Springer), 421–446. doi: 10.1007/978-1-0716-2197-4_11

Kroese, F., Marchiori, D., and de Ridder, D. (2015). Nudging healthy food choices: a field experiment at the train station. *J. Publ. Health* 38, e133–e137. doi: 10.1093/pubmed/fdv096

Lannelongue, L., Aronson, H., Bateman, A., Birney, E., Caplan, T., Juckes, M., et al. (2023). GREENER principles for environmentally sustainable computational science. *Nat. Comput. Sci.* 3, 514–521. doi: 10.1038/s43588-023-00461-y

Lehner, M., Mont, O., and Heiskanen, E. (2016). Nudging-a promising tool for sustainable consumption behaviour? *J. Clean. Prod.* 134, 166–177. doi: 10.1016/j.jclepro.2015.11.086

Lex, E., Kowald, D., Seitlinger, P., Tran, T., Felfernig, A., and Schedl, M. (2021). Psychology-informed recommender systems. *Found. Trends Inform. Retriev.* 15, 134–242. doi: 10.1561/1500000090

Li, M., Bao, X., Chang, L., Xu, Z., and Li, L. (2020). "A survey of researches on personalized bundle recommendation techniques," in *3rd International Conference on Machine Learning for Cyber Security* (Cham: Springer), 290–304. doi: 10.1007/978-3-030-62460-6_26

Li, Y., Chen, H., Xu, S., Ge, Y., Tan, J., Liu, S., et al. (2023). Fairness in recommendation: foundations, methods and applications. *ACM Trans. Intell. Syst. Technol.* 14, 1–46. doi: 10.1145/3610302

Liang, Z. (2022). Context-aware sleep health recommender systems (CASHRS): a narrative review. *Electronics* 11:3384. doi: 10.3390/electronics11203384

Liao, H., Huang, X.-M., Wu, X.-T., Liu, M.-K., Vidmer, A., Zhou, M.-Y., et al. (2018). Enhancing Countries' fitness with recommender systems on the international trade network. *Complexity* 2018:5806827. doi: 10.1155/2018/5806827

Luef, J., Ohrfandl, C., Sacharidis, D., and Werthner, H. (2020). "A recommender system for investing in early-stage enterprises," in *Proceedings of the 35th Annual ACM Symposium on Applied Computing, SAC '20* (New York, NY), 1453–1460. doi: 10.1145/3341105.3375767

Magalhães, F., de Queiroz, A., Machado, B., and Paulo, P. (2019). Sustainable sanitation management tool for decision making in isolated areas in Brazil. *Int. J. Environ. Res. Publ. Health* 16:1118. doi: 10.3390/ijerph16071118

Mahmoud, S., El-Bendary, N., Mahmood, M., and Hassanien, A. (2013). "An intelligent recommender system for drinking water quality," in *13th International Conference on Hybrid Intelligent Systems (HIS 2013)* (New York, NY), 285–290. doi: 10.1109/HIS.2013.6920498

Majjodi, A. E., Elahi, M., Ioini, N. E., and Trattner, C. (2020). "Towards generating personalized country recommendation," in *28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '20 Adjunct* (New York, NY), 71–76. doi: 10.1145/3386392.3397601

Majjodi, A. E., Starke, A., and Trattner, C. (2022). "Nudging towards health? Examining the merits of nutrition labels and personalization in a recipe recommender system," in *30th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '22* (New York, NY), 48–56. doi: 10.1145/3503252.3531312

Mandal, A., Chaki, R., Saha, S., Pal, K. G. A., and Ghosh, S. (2017). "Measuring similarity among legal court case documents," in *10th Annual ACM India Compute Conference, Compute '17* (New York, NY), 1–9. doi: 10.1145/3140107.3140119

Mandl, M., Felfernig, A., Teppan, E., and Schubert, M. (2011). Consumer decision making in knowledge-based recommendation. *J. Intell. Inform. Syst.* 37, 1–22. doi: 10.1007/s10844-010-0134-3

Martini, G., Bracci, A., Riches, L., Jaiswal, S., Corea, M., Rivers, J., et al. (2022). Machine learning can guide food security efforts when primary data are not available. *Nat. Food* 3, 716–728. doi: 10.1038/s43016-022-00587-8

Mauro, N., Ardissono, L., and Cena, F. (2022). Supporting people with autism spectrum disorders in the exploration of Pois: an inclusive recommender system. *Commun. ACM* 65, 101–109. doi: 10.1145/3505267

Merinov, P. (2023). "Sustainability-oriented recommender Systems," in *31st ACM Conference on User Modeling, Adaptation and Personalization, UMAP '23* (New York, NY), 296–300. doi: 10.1145/3565472.3595617

Muralidhar, R., Borovica-Gajic, R., and Buyya, R. (2022). Energy efficient computing systems: architectures, abstractions and modeling to techniques and standards. *ACM Comput. Surveys* 54, 1–37. doi: 10.1145/3511094

Murphy, S., Manzano, O., and Brown, K. (2015). "Design and evaluation of a constraint-based energy saving and scheduling recommender system," in *Principles and Practice of Constraint Programming (CP 2015), Vol. 9255 of Lecture Notes in Computer Science* (Cham: Springer), 687–703. doi: 10.1007/978-3-319-23219-5_47

Nguyen, T., Sinha, A., Gholami, S., Plumptre, A., Joppa, L., Tambeand, M., et al. (2016). "CAPTURE: a new predictive anti-poaching tool for wildlife protection," in *International Conference on Autonomous Agents & Multiagent Systems, AAMAS '16* (Richland, SC), 767–775.

Omara, J., Talavera, E., Otim, D., Turcza, D., Ofumbi, E., and Owomugisha, G. (2023). A field-based recommender system for crop disease detection using machine learning. *Front. Artif. Intell.* 6:1010804. doi: 10.3389/frai.2023.1010804

Patel, K., and Patel, H. (2020). A state-of-the-art survey on recommendation system and prospective extensions. *Comput. Electron. Agric.* 178:105779. doi: 10.1016/j.compag.2020.105779

Pazzani, M., and Billsus, D. (2007). "Content-based recommendation systems," in *The Adaptive Web: Methods and Strategies of Web Personalization* (Berlin: Springer), 325–341. doi: 10.1007/978-3-540-72079-9_10

Petkov, P., Köbler, F., Foth, M., and Krcmar, H. (2011). "Motivating domestic energy conservation through comparative, community-based feedback in mobile and social media," in *5th International Conference on Communities and Technologies, C&T '11* (New York, NY), 21–30. doi: 10.1145/2103354.2103358

Pinciroli, L., Baraldi, P., Ballabio, G., Compare, M., and Zio, E. (2022). Optimization of the operation and maintenance of renewable energy systems by deep reinforcement learning. *Renew. Energy* 183, 752–763. doi: 10.1016/j.renene.2021.11.052

Praba, M., Roy, S., Reddy, M., and Yadav, I. (2023). "Smart fish farming recommendation system using k-means algorithm," in *13th International Conference on Cloud Computing, Data Science & Engineering* (New York, NY), 333–338. doi: 10.1109/Confluence56041.2023.10048806

Quisi-Peralta, D., Chiluisa-Castillo, D., Robles-Bykbaev, V., López-Nores, M., and Chaglla-Rodriguez, L. (2018). "A text filter based multimedia content recommender for children with intellectual disability," in *25th International Conference on Electronics, Electrical Engineering and Computing (INTERCON)* (New York, NY), 1–4. doi: 10.1109/INTERCON.2018.8526440

Ravanmehr, R. Z. R. (2021). Serendipity in recommender systems: a systematic literature review. *J. Comput. Sci. Technol.* 36, 375–396. doi: 10.1007/s11390-020-0135-9

Ribeiro, A. (2011). "A model of joint learning in poverty: coordination and recommendation systems in low-income communities," in *4th International Conference on Machine Learning and Applications* (New York, NY), 63–67. doi: 10.1109/ICMLA.2011.15

Roher, K., and Richardson, D. (2013). "A proposed recommender system for eliciting software sustainability requirements," in *2nd International Workshop on User Evaluations for Software Engineering Researchers*, 16–19. doi: 10.1109/USER.2013.6603080

Ronzhin, A., Ngo, T., Vu, O., and Nguyen, V. (2022). "Recommendation system to select the composition of the heterogeneous agricultural robots," in *Ground and Air Robotic Manipulation Systems in Agriculture* (Cham: Springer), 45–63. doi: 10.1007/978-3-030-86826-0_3

Sallami, D., Ben, S., and Aïmeur, E. (2023). "Trust-based recommender system for fake news mitigation," in *31st ACM Conference on User Modeling, Adaptation and Personalization, UMAP '23 Adjunct* (New York, NY), 104–109. doi: 10.1145/3563359.3597395

Shadowen, N., Lodato, T., and Loi, D. (2020). "Participatory governance in smart cities: future scenarios and opportunities," in *8th International Conference on Distributed, Ambient and Pervasive Interactions* (Cham: Springer), 443–463. doi: 10.1007/978-3-030-50344-4_32

Shi, Z., Lizarondo, L., and Fang, F. (2021). "A recommender system for crowdsourcing food rescue platforms," in *Web Conference 2021* (New York, NY), 857–865. doi: 10.1145/3442381.3449787

Sihotang, D., Hidayanto, A., Abidin, Z., and Diana, E. (2021). Smart method for recommender system towards smart tourism and green computing. *IOP Conf. Ser.* 700:012017. doi: 10.1088/1755-1315/700/1/012017

Singh, H., Singh, M. B., Sharma, R., Gat, J., Agrawal, A. K., and Pratap, A. (2023). "Optimized doctor recommendation system using supervised machine learning," in *24th International Conference on Distributed Computing and Networking* (New York, NY), 360–365. doi: 10.1145/3571306.3571372

Smith, B., and Linden, G. (2017). Two decades of recommender systems at Amazon.Com. *IEEE Intern. Comput.* 21, 12–18. doi: 10.1109/MIC.2017.72

Smith, R., and Iversen, O. (2018). Participatory design for sustainable social change. *Des. Stud.* 59, 9–36. doi: 10.1016/j.destud.2018.05.005

Sonboli, N., Burke, R., Ekstrand, M., and Mehrotra, R. (2022). the multisided complexity of fairness in recommender systems. *AI Mag.* 43, 164–176. doi: 10.1002/aaai.12054

Spillo, G., Filippo, A. D., Musto, C., Milano, M., and Semeraro, G. (2023). "Towards sustainability-aware recommender systems: analyzing the trade-off between algorithms performance and carbon footprint," in *17th ACM Conference on Recommender Systems, RecSys '23* (New York, NY), 856–862. doi: 10.1145/3604915.3608840

Starke, A., Willemsen, M., and Snijders, C. (2017). "Effective user interface designs to increase energy-efficient behavior in a Rasch-based energy recommender system," in *11th ACM Conference on Recommender Systems, RecSys '17* (New York, NY), 65–73. doi: 10.1145/3109859.3109902

Starke, A., Willemsen, M., and Snijders, C. (2021). "Using explanations as energy-saving frames: a user-centric recommender study," in *29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '21 Adjunct* (New York, NY), 229–237. doi: 10.1145/3450614.3464477

Stettinger, M., Tran, T., Pribik, I., Leitner, G., Felfernig, A., Samer, R., et al. (2020). "KnowledgeCheckR: intelligent techniques for counteracting forgetting," in *24th European Conference on Artificial Intelligence (ECAI 2020)* (Amsterdam: IOS Press), 3034–3039.

Stray, J., Vendrov, I., Nixon, J., Adler, S., and Hadfield-Menell, D. (2021). What are you optimizing for? Aligning recommender systems with human values. *ArXiv:abs/2107.10939*. doi: 10.48550/arXiv.2107.10939

Sultana, U., Maha, A., Maha, I., and Midhat, F. (2022). Increase the performance of wind energy systems using optimal layout planning. *Eng. Proc.* 20:20. doi: 10.3390/engproc2022020020

Tavakoli, M., Faraji, A., Vrolijk, J., Molavi, M., Mol, S., and Kismihók, G. (2022). An AI-based open recommender system for personalized labor market driven education. *Adv. Eng. Inform.* 52:101508. doi: 10.1016/j.aei.2021.101508

Tayebi, M., Jamali, M., Ester, M., Glässer, U., and Frank, R. (2011). "CrimeWalker: a recommendation model for suspect investigation," in *5th ACM Conference on Recommender Systems, RecSys '11* (New York, NY), 173–180. doi: 10.1145/2043932.2043965

Thaler, R., and Sunstein, C. (2021). *Nudge: The Final Edition*. London: Penguin Books.

Thilakarathne, N., Bakar, M., Abas, P., and Yassin, H. (2022). A cloud enabled crop recommendation platform for machine learning-driven precision farming. *Sensors* 22:16. doi: 10.3390/s22166299

Tintarev, N., and Masthoff, J. (2012). Evaluating the effectiveness of explanations for recommender systems: methodological issues and empirical studies on the impact of personalization. *User Model. User Adapt. Interact.* 22, 399–439. doi: 10.1007/s11257-011-9117-5

Tomita, Y., Togashi, R., and Moriwaki, D. (2022). "Matching theory-based recommender systems in online dating," in *16th ACM Conference on Recommender Systems, RecSys '22* (New York, NY), 538–541. doi: 10.1145/3523227.3547406

Tomkins, S., Isley, S., London, B., and Getoor, L. (2018). "Sustainability at scale: towards bridging the intention-behavior gap with sustainable recommendations," in *12th ACM Conference on Recommender Systems, RecSys '18* (New York, NY), 214–218. doi: 10.1145/3240323.3240411

Tran, T., Atas, M., Felfernig, A., and Stettinger, M. (2018a). An overview of recommender systems in the healthy food domain. *J. Intell. Inform. Syst.* 50, 501–526. doi: 10.1007/s10844-017-0469-0

Tran, T., Felfernig, A., Trattner, C., and Holzinger, A. (2018b). Recommender systems in the healthcare domain: state-of-the-art and research issues. *J. Intell. Inform. Syst.* 57, 171–201. doi: 10.1007/s10844-020-00633-6

Tran, T. N. T., Felfernig, A., Le, V. M., Atas, M., Stettinger, M., and Samer, R. (2019). "User interfaces for counteracting decision manipulation in group recommender systems," in *27th Conference on User Modeling, Adaptation and Personalization, UMAP'19 Adjunct* (New York, NY), 93–98. doi: 10.1145/3314183.3324977

Tran, T. N. T., Felfernig, A., and Tintarev, N. (2021). Humanized recommender systems: state-of-the-art and research issues. *ACM Trans. Interact. Intell. Syst.* 11, 1–41. doi: 10.1145/3446906

Tsai, K., Yang, F., and Tang, C. (2022). Multiagent mobility and lifestyle recommender system for individuals with visual impairment. *Neurosci. Inform.* 2:100077. doi: 10.1016/j.neuri.2022.100077

Tversky, A., and Kahneman, D. (1985). "The framing of decisions and the psychology of choice," in *Behavioral Decision Making*, ed G. Wright (Boston, MA: Springer), 25–41. doi: 10.1007/978-1-4613-2391-4_2

Usman, A., Lin, J., Srivastava, G., and Y, D. (2021). A nutrient recommendation system for soil fertilization based on evolutionary computation. *Comput. Electron. Agric.* 189:106407. doi: 10.1016/j.compag.2021.106407

Usmanova, A., Aziz, A., Rakhmonov, D., and Osamy, W. (2022). Utilities of artificial intelligence in poverty prediction: a review. *Sustainability* 14:21. doi: 10.3390/su142114238

Vaghasiya, T., Bhatt, N., Vallee, F., Lecron, F., and de Greve, Z. (2017). Wind power prediction by using matrix factorization technique. *Int. J. Adv. Eng. Res. Dev.* 4, 351–362. doi: 10.21090/IJAERD.47743

van Capelleveen, G., Amrit, C., Yazan, D., and Zijm, H. (2018). The influence of knowledge in the design of a recommender system to facilitate industrial symbiosis markets. *Environ. Model. Softw.* 110, 139–152. doi: 10.1016/j.envsoft.2018.04.004

vanWynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI Ethics* 1, 213–218. doi: 10.1007/s43681-021-00043-6

Vidal-Silva, C., Galindo, J., Giráldez-Cru, J., and Benavides, D. (2021). "Automated completion of partial configurations as a diagnosis task using FastDiag to improve performance," in *Intelligent Systems in Industrial Applications, Vol. 949* (Cham), 107–117. doi: 10.1007/978-3-030-67148-8_9

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., et al. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nat. Commun.* 1:223. doi: 10.1038/s41467-019-14108-y

Wakchaure, M., Patle, B., and Mahindrakar, A. (2023). Application of AI techniques and robotics in agriculture: a review. *Artif. Intell. Life Sci.* 3:100057. doi: 10.1016/j.ailsci.2023.100057

Wang, W., Duan, L.-Y., Jiang, H., Jing, P., Song, X., and Nie, L. (2021). Market2Dish: health-aware food recommendation. *ACM Trans. Multim. Comput. Commun. Appl.* 17, 1–19. doi: 10.1145/3418211

Wei, P., Xia, S., and Jiang, X. (2018). "Energy saving recommendations and user location modeling in commercial buildings," in *26th Conference on User Modeling, Adaptation and Personalization, UMAP '18* (New York, NY), 3–11. doi: 10.1145/3209219.3209244

Wiezorek, R., and Christensen, N. (2021). "Integrating sustainability information in configurators," in *Workshop on Configuration, Vol. 2945 of ConfWS'21*, 65–72.

Wohlin, C. (2014). "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *18th International Conference on Evaluation and Assessment in Software Engineering* (New York, NY), 1–10. doi: 10.1145/2601248.2601268

Wu, Y., Cao, J., and Xu, G. (2023). Fairness in recommender systems: evaluation approaches and assurance strategies. *ACM Trans. Knowl. Discov. Data* 18, 1–37. doi: 10.1145/3604558

Wu, Y., Macdonald, C., and Ounis, I. (2022). "Multimodal conversational fashion recommendation with positive and negative natural-language feedback," in *4th Conference on Conversational User Interfaces, CUI '22* (New York, NY), 1–10. doi: 10.1145/3543829.3543837

Xu, X., Lai, T., Jahan, S., Farid, F., and Bello, A. (2022). a machine learning predictive model to detect water quality and pollution. *Fut. Intern.* 14:324. doi: 10.3390/fi14110324

Yang, R., Ford, B., Tambe, M., and Lemieux, A. (2014). "Adaptive resource allocation for wildlife protection against illegal poachers," in *International Conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14* (Richland, SC), 453–460.

Zangerle, E., and Bauer, C. (2022). Evaluating recommender systems: survey and framework. *ACM Comput. Surv.* 55, 1–38. doi: 10.1145/3556536

Zielnicki, K. (2019). "Simulacra and selection: clothing set recommendation at stitch fix," in *42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'19* (New York, NY), 1379–1380. doi: 10.1145/3331184.3331442

frontiers | Frontiers in Big Data

TYPE Review
PUBLISHED 30 October 2023
DOI 10.3389/fdata.2023.1281614

Check for updates

OPEN ACCESS

EDITED BY
Emanuel Lacić,
Infobip d.o.o., Croatia

REVIEWED BY
Pasquale Lops,
University of Bari Aldo Moro, Italy
Junyang Chen,
Shenzhen University, China

*CORRESPONDENCE
Sebastian Lubos
slubos@ist.tugraz.at

RECEIVED 22 August 2023
ACCEPTED 06 October 2023
PUBLISHED 30 October 2023

CITATION
Lubos S, Felfernig A and Tautschnig M (2023) An overview of video recommender systems: state-of-the-art and research issues.
Front. Big Data 6:1281614.
doi: 10.3389/fdata.2023.1281614

COPYRIGHT
© 2023 Lubos, Felfernig and Tautschnig. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An overview of video recommender systems: state-of-the-art and research issues

Sebastian Lubos[1]*, Alexander Felfernig[1] and Markus Tautschnig[2]

[1]Applied Software Engineering & AI Research Group, Institute of Software Technology, Graz University of Technology, Graz, Austria, [2]Streamdiver GmbH, Klagenfurt am Wörthersee, Austria

Video platforms have become indispensable components within a diverse range of applications, serving various purposes in entertainment, e-learning, corporate training, online documentation, and news provision. As the volume and complexity of video content continue to grow, the need for personalized access features becomes an inevitable requirement to ensure efficient content consumption. To address this need, recommender systems have emerged as helpful tools providing personalized video access. By leveraging past user-specific video consumption data and the preferences of similar users, these systems excel in recommending videos that are highly relevant to individual users. This article presents a comprehensive overview of the current state of *video recommender systems (VRS)*, exploring the algorithms used, their applications, and related aspects. In addition to an in-depth analysis of existing approaches, this review also addresses unresolved research challenges within this domain. These unexplored areas offer exciting opportunities for advancements and innovations, aiming to enhance the accuracy and effectiveness of personalized video recommendations. Overall, this article serves as a valuable resource for researchers, practitioners, and stakeholders in the video domain. It offers insights into cutting-edge algorithms, successful applications, and areas that merit further exploration to advance the field of video recommendation.

KEYWORDS

video recommender systems, collaborative filtering, content-based recommendation, hybrid recommenders, group recommenders, decision-making, overview, research challenges

1. Introduction

Recommender systems (RS) support various decision-making scenarios ranging from the recommendation of simple items, such as books or movies, to more complex ones, like financial services and digital equipment (Ricci et al., 2011). Among these applications, *movie recommender systems* stand out as a pioneering example, suggesting movies that users may find interesting to watch (Harper and Konstan, 2015). These movie recommenders are a specific category within *video recommender systems (VRS)*, which are gaining significant attention in entertainment, as well as industrial contexts, due to the rapidly increasing number of available video items.

Popular video platforms, for example, YouTube[1] and Netflix,[2] integrate recommendation technologies to enhance user experience by suggesting videos from their huge catalogs that are likely to align with users' personal interests and preferences (Davidson et al., 2010; Gomez-Uribe and Hunt, 2016). From an economic perspective, these platforms aim to attract and retain customers, increasing the retention rate through effective content recommendations (Gomez-Uribe and Hunt, 2016). For instance, around two-thirds of the content streamed on Netflix originates from recommendations featured on the entry page (Gomez-Uribe and Hunt, 2016). Moreover, empirical studies have demonstrated that video recommendations can capture a user's attention toward specific topics and consequently increase the popularity of particular videos (Wu et al., 2019), emphasizing the power of this technology.

Several reviews related to video recommendations have been published in the past years. In Véras et al. (2015) recommender systems in the television domain are covered, including content related to TV shows. In Wang and Zhao (2022), an in-depth analysis of affective video recommender systems, i.e., systems that integrate human-like capabilities of observation, interpretation, and generation of affect features, like, emotions and mood, is provided. A broader overview of multimedia item recommenders, encompassing audio, images, and videos, is presented in Deldjoo et al. (2022), focusing on methods for feature extraction and integration of multimedia data as side information in recommenders. In Jayalakshmi et al. (2022), a literature review on movie recommender systems is provided, discussing algorithmic commonalities and recent publications in this domain.

While those related reviews specialize in specific video-related recommender aspects, our overview provides a concise summary of video item recommendations, serving as a comprehensible summary of the state-of-the-art for practitioners and researchers in this area. This overview should enhance understanding of the various technical approaches within this field and their applications. Additionally, it identifies open issues that should be addressed in future research to further develop the field.

The article is structured as follows: In Section 2, we outline the analysis method employed in our literature review. In Section 3, we conduct an in-depth analysis of the existing literature on VRS, categorizing it based on different fundamental approaches of recommender systems and the technologies utilized. Following that, in Section 4, we discuss the findings and offer insights to comprehend which approaches excel in various recommendation scenarios. Additionally, we address future research considerations and discuss unresolved issues. Finally, the article concludes in Section 5.

The major contributions of this article can be summarized as follows: *Firstly*, we present an extensive overview of the current state-of-the-art in VRS, covering research developments from the past decades. *Secondly*, we provide valuable guidance for selecting suitable recommendation approaches based on individual scenarios. *Thirdly*, we engage in a comprehensive discussion of open research issues, highlighting the potential for future work in this evolving field of research.

1 www.youtube.com
2 www.netflix.com

2. Methods

The main objective of this article is to provide an overview of state-of-the-art video recommender systems to increase understanding of this topic, derive guidance for choosing appropriate approaches, and identify issues for future research. In this context, we include recommender systems where the recommended items are *videos*, independent of the domain. This includes entertainment, e.g., movies or videos on social networks, as well as video advertisements, learning videos, news videos, and others.

Our analysis of related work is based on a bibliographic review method. As an initial step, we collected and reviewed existing publications on VRS over the last 20 years. The search for related papers was performed on the basis of different keywords, including, *"video recommender systems"*, *"video recommender"*, *"video recommendation"*, *"movie recommender systems"*, *"movie recommender"*, and *"movie recommendation"*. With these, queries were triggered in the digital libraries of ACM,[3] Google Scholar,[4] ResearchGate,[5] Science Direct,[6] and Springer Link.[7]

Following the review, publications were categorized by their recommendation approach (content-based, collaborative, hybrid, and group recommendation), and further divided into subcategories of different applied algorithms. The results are outlined below. The topic of video content representation, which is relevant for content-based and many hybrid recommender approaches, is summarized in a separate section. From these findings, guidance in selecting appropriate technologies is derived and open topics for future research are identified.

3. Video recommender systems

Video recommender systems suggest videos to users based on their individual preferences. An overview of a typical pipeline used for video recommendation is illustrated in Figure 1. A specialty for recommendations in the video domain is the representation of content in terms of features that are automatically extracted or manually added. Videos offer a rich variety of different features that can be used to describe their content. Details on content representation are discussed in Section 3.1.

Similar to recommendations in other item domains, dealing with a large catalog of videos can lead to performance issues. To address this, a common approach is to split the computation in a *retrieval* and *ranking* phase (Davidson et al., 2010; Covington et al., 2016; Gomez-Uribe and Hunt, 2016). The retrieval phase reduces the number of candidates to a reasonable number using

3 https://dl.acm.org
4 https://scholar.google.com
5 https://www.researchgate.net
6 https://www.sciencedirect.com
7 https://link.springer.com

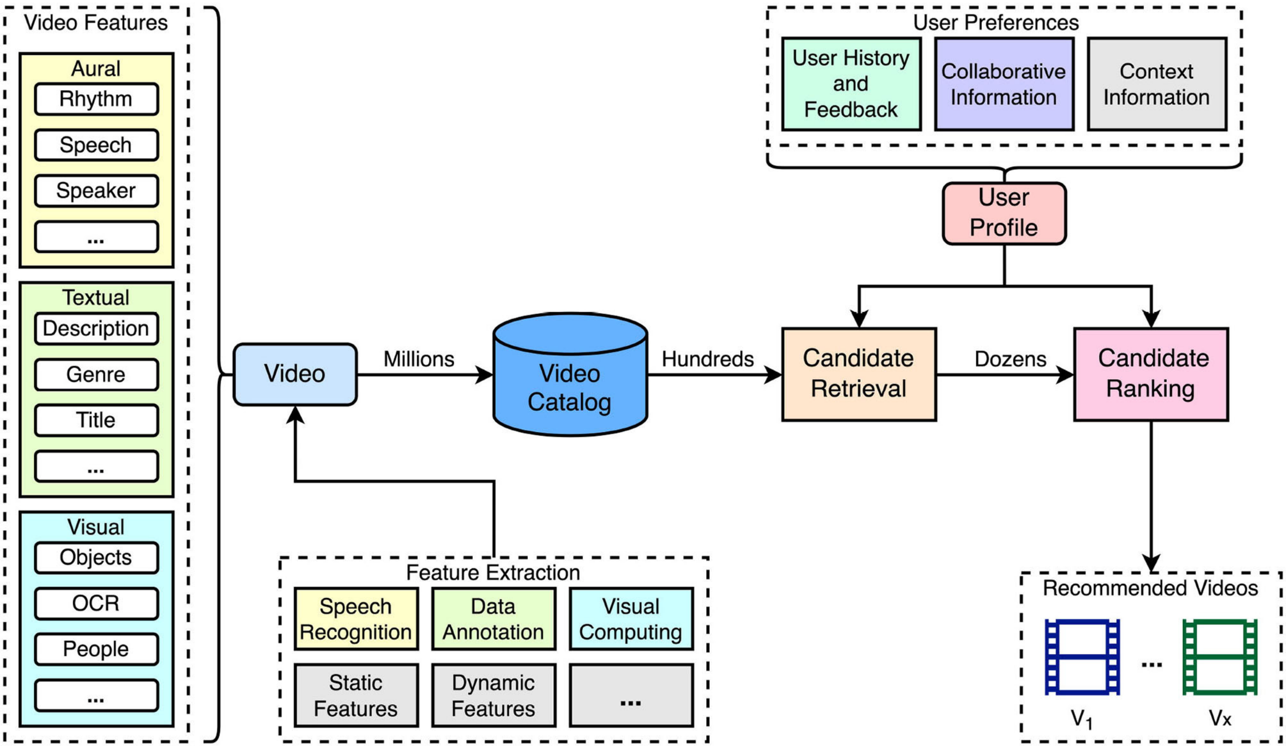

FIGURE 1
Overview of the pipeline used in video recommender systems. Typically, videos are indexed in the catalog using feature descriptions that are either automatically extracted or added manually (see Section 3.1). Using the videos in the catalog, personalized recommendations are retrieved in a two-step phase by identifying candidates and ranking them based on the generated user profile describing their preferences.

a relatively fast analysis. In the ranking phase, the remaining candidates are ordered by relevance using more precise but often slower algorithms. This two-step strategy enables efficient video recommendations from extensive catalogs within an acceptable time. Both steps consider a user profile generated from information, such as the user history of consumed videos, provided feedback, information of similar users, and the current user context.

The variability of VRS applications can be illustrated by taking NETFLIX as an example (Gomez-Uribe and Hunt, 2016). The platform uses a *personalized video ranker (PVR)* algorithm to order its video catalog based on user profiles, video popularity, and temporal viewing trends. Different algorithms are applied on top for various purposes: (1) Identifying the most relevant items from the catalog for each user. (2) Ordering videos users have started watching. (3) Unpersonalized prediction of short-term temporal trends for events like Halloween or Christmas, or unplanned incidents, such as a hurricane or other natural catastrophes currently populated by the news. (4) Recommending videos with similar content. (5) Enhancing content presentation by selecting thumbnails and presented metadata. Furthermore, NETFLIX employs a *page generation* algorithm to define the selection and ordering of rows presented in the UI. It considers that one account is mostly used by multiple users, e.g., family members, aiming for a diverse content presentation that is relevant to each user in front of the screen.

In the following, the literature on VRS is discussed. Foremost, the methods used to represent the content of videos are discussed. Subsequently, publications are categorized by the applied recommendation approach, including *content-based recommendation*, *collaborative filtering (CF)*, *hybrid recommendation*, and *group recommendation*. In Figure 2, a simplified overview of the different approaches is shown. While content-based recommendation (see Figure 2A) recommends videos to a user based on their similarity, collaborative filtering (see Figure 2B) exploits the knowledge of users with similar interests. Hybrid recommenders (see Figure 2C) combine different approaches to generate recommendations. While the aforementioned approaches focus on recommending items to individual users, group recommenders (see Figure 2D) try to suggest videos that are in line with the preferences of a group consisting of multiple persons.

3.1. Content representation

Video recommenders differ notably from those in many other domains, e.g., shopping, due to the nature of the items being recommended. Unlike structured features like color, brand, category, or price that describe shopping items, video content descriptions encompass more possibilities due to their *multimodality*. Videos consist of three modalities: (1) *Aural* (audio information), (2) *Visual* (visual frames), and (3) *Textual* (textual descriptions and metadata), which can be expressed in varying

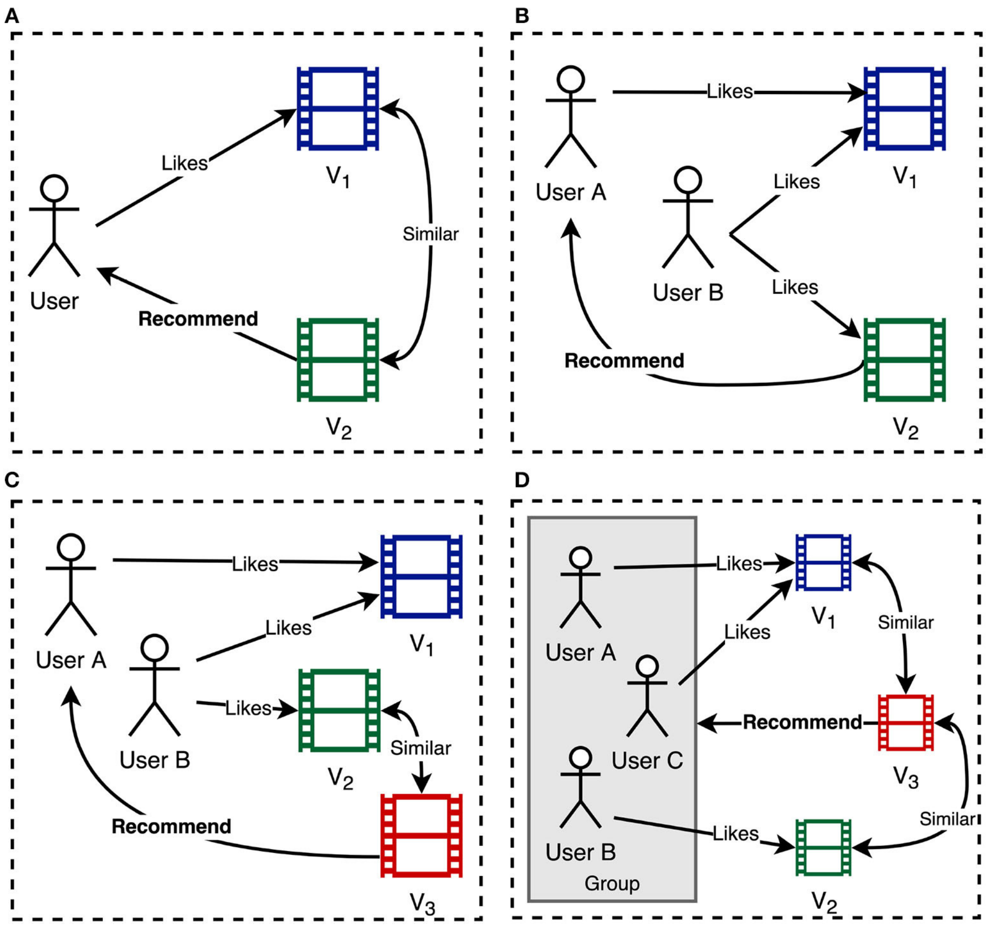

FIGURE 2
Simplified overview of different recommendation approaches. **(A)** Content-based recommendation. **(B)** Collaborative filtering. **(C)** Hybrid recommendation. **(D)** Group recommendation.

degrees of semantic detail. This characteristic makes videos *multi-modal*, as they include all three modalities, whereas a music piece without lyrics is *uni-modal*, as it only features aural elements (Deldjoo, 2020).

Based on the classification in Deldjoo (2020), video features can be categorized into groups based on their modality and semantic expressiveness: (1) *Low-level* features describe the raw signal of a video, representing its stylistic properties. (2) *Mid-level* features require interpretation knowledge and are derived from low-level features, representing syntactic features. (3) *High-level* features resemble human interpretation of the content, providing a semantic description. Table 1 presents an overview of these categorized features, enabling the classification of VRS based on the features they use for computing recommendations.

Content descriptions in the video domain can be manually created or automatically extracted. Manual features typically include a title, a short description, and tags. For movies, databases like *Internet Movie Database (IMDb)*[8] and *Open Movie Database (OMDB)*[9] provide structured metadata including actors, genres, plots, and more. Another option is the extension with *semantic web data*, illustrated in Hopfgartner and Jose (2010), which leverages Linked Open Data Cloud[10] for content description enrichment.

Automatically extracted features in the video domain offer diverse options in semantic expressiveness and modalities. A common technique is the conversion into *embeddings*, representing words as numerical vectors in a lower-dimensional space, preserving item feature information (Huang et al., 2019). This approach provides a compact representation and enables mathematical operations on the embeddings.

Videos share similarities in processing with other multimedia items like audio and images. Image processing methods can be applied to video frames for visual feature retrieval, while audio processing techniques analyze the audio track. Yet, videos offer additional temporal attributes, enabling action and motion recognition over time. For more details on fundamental extraction methods for multimedia items, refer to Deldjoo et al. (2022). In this overview, we focus on algorithmic approaches and applications of video recommenders, utilizing both manually created and automatically extracted features.

8 https://www.imdb.com/
9 http://www.omdbapi.com/
10 https://lod-cloud.net

TABLE 1 Multimedia features categorized by their expressiveness and modality.

Hierarchy	Modalities		
	Aural	Textual	Visual
Low-level	Beat, frequency, loudness, intensity, pitch, timbre	n-grams, tokens	Colors, contours, edges, key points, keyframes, motions, shapes, textures
Mid-level	Note onsets, patterns, rhythm, tempo	Paragraphs, sentences, term-frequencies, transcript	Actions, interactions, objects, people, scenes, shots, scenes
High-level	Events, mood, speech, speaker, story	Comments, description, genre, events, keywords, key phrases, named entities, sentiment, story, tags, title, topic, writing style	Concept, emotion, message, language, speaker, structure

The table extends the one provided in Deldjoo (2020).

Table 2 summarizes the content modalities used in various video domains. The table shows that research on VRS with content representations predominantly focuses on *Movies and Series* and videos within *Social Networks*. Reasons might be the significant user base and availability of datasets in these domains (see Section 3.6.4). Initiatives like the *Netflix Prize* have contributed to this emphasis by providing real-life data to improve movie recommendation accuracy (Bennett and Lanning, 2007).

Based on the summary, video recommender system research has employed diverse modalities to represent video content, revealing certain trends. Aural features were infrequently used, and when applied, were often combined with textual or visual features. This implies that sole reliance on aural features might lack accuracy. Visual features were prevalent, especially in entertainment domains, where visuals are significant. Textual features were widely adopted across domains, likely due to the ability to reuse technical approaches from other domains and the rich information they provide, particularly in educational videos where facts are more relevant than visual aspects.

Generating appropriate suitable video representations is crucial in video recommendation and has been extensively studied. The study in Elahi et al. (2017) focused on the *semantic gap*, which refers to the difference between various representations of the same item. The study evaluated various video representations and found that both low-level stylistic features (e.g., brightness and contrast) and high-level semantic concepts (e.g., genre and actors) contribute to accurate recommendations. Combining these features through a multi-modal approach showed potential for improving accuracy.

A related study found similar results with automatically extracted aural and visual features (Deldjoo et al., 2018a). Aural features included short audio segment characteristics (*Block-Level-Features*) and low-dimensional representations of acoustic signals (*I-Vector Features*). Visual features included *Aesthetic Visual Features (AVF)*, categorized by color, texture, and objects, as well as high-level features extracted with *Deep Neural Networks (DNN)*. Utilizing multi-modal representations with weighted aggregation again demonstrated the potential for improving accuracy.

The positive impact of multi-modal representations with automatically extracted aural and visual features was also observed in Lee and Abu-El-Haija (2017), where optimization options for embedding representations were explored. Increasing the output feature size of embeddings, utilizing deeper models, enhancing the capacity of the first hidden layer, and applying late fusion of aural and visual features led to more accurate recommendations. The representations were found to capture the semantic features of items, despite the features themselves not being inherently semantic. Moreover, the representations proved effective in accurately recommending videos on the same topic but in different languages. The possibility to predict descriptive tags for videos from low-level visual features was described (Elahi et al., 2020), confirming the possibility to generate features with semantic meaning from unsemantic data.

In Pingali et al. (2022), a multi-modal content representation approach for movies is proposed, which involves concatenating feature embeddings from aural and visual features, textual descriptions, and other metadata to create a vector representation of the video in a vector space. Those unsupervised methods for generating content representations help address the challenge of cold start, where limited or no initial information is available and reduce manual effort at the same time (Hazrati and Elahi, 2021).

The study in Deldjoo et al. (2016) highlights essential findings regarding video representation. Low-level visual features from movie trailers accurately capture the full movie's essence, enabling performance tuning with smaller samples. Automatic extraction of visual features addresses missing content descriptions for competitive accuracy in recommendations. However, combining various features might reduce accuracy due to a lack of correlation between them. Subsequent research in Deldjoo et al. (2018b) validates this, showing that maximizing pairwise correlation through feature fusion does not enhance accuracy, suggesting that stylistically similar movies might not share semantic commonalities.

Each visual feature has different capabilities to capture the video content appropriately and thus can contribute differently to the creation of recommendations (Hazrati and Elahi, 2021). Combining features can enhance recommendation accuracy if their information is not contradicting. The same is true for the aural features of videos (Rimaz et al., 2021).

High-level visual features such as faces, objects, and recognized celebrities were automatically extracted in Elahi et al. (2021), to create vector representations for videos using a combination of *term frequency-inverse document frequency (TF-IDF)* (Sammut and Webb, 2010) and *word2vec* (Mikolov et al., 2013). TF-IDF is a statistical measure that reflects the importance of terms within a document or catalog, while word2vec describes a DNN technique used in *Natural Language Processing (NLP)* to learn word relationships. This representation incorporating semantic features allows for human comprehension

TABLE 2 Content representations in VRS classified by domain and used feature modalities.

Domain	Features	References
Advertisement	Textual	Kaklauskas et al., 2018; Kim et al., 2021
Education	Textual	Chantanurak et al., 2016; Kimoto et al., 2016; Tavakoli et al., 2020; Leite et al., 2022
Movies and Series	Aural	Deldjoo et al., 2018a; Rimaz et al., 2021; Chakder et al., 2022; Pingali et al., 2022; Mondal et al., 2023
	Textual	Öztürk and Kesim Cicekli, 2011; Zhu et al., 2013; Vizine Pereira and Hruschka, 2015; Wang et al., 2015, 2021; Gomez-Uribe and Hunt, 2016; Elahi et al., 2017; Lu et al., 2017; Wei et al., 2017; Liu et al., 2019b; Kvifte et al., 2021; Zhuo et al., 2021; Chakder et al., 2022; Pingali et al., 2022; Mondal et al., 2023
	Visual	Zhu et al., 2013; Deldjoo et al., 2016, 2018a,b; Elahi et al., 2017, 2020, 2021; Hazrati and Elahi, 2021; Kvifte et al., 2021; Wang et al., 2021; Chakder et al., 2022; Pingali et al., 2022; Mondal et al., 2023
News	Aural	Luo et al., 2008
	Textual	Luo et al., 2008; Hopfgartner and Jose, 2010
	Visual	Luo et al., 2008
Social Networks	Aural	Mei et al., 2007, 2011; Niu et al., 2013; Lee and Abu-El-Haija, 2017; Liu et al., 2019a; Du et al., 2022; Yi et al., 2022
	Textual	Mei et al., 2007, 2011; Wu et al., 2008; Davidson et al., 2010; Cui et al., 2014; Covington et al., 2016; Abbas et al., 2017; Gao et al., 2017; Chen et al., 2018, 2021; Li et al., 2019; Liu et al., 2019a; Jiang et al., 2020; Tang et al., 2020; Du et al., 2022; Gong et al., 2022; Yi et al., 2022; Song et al., 2023; Xiao et al., 2023
	Visual	Mei et al., 2007, 2011; Niu et al., 2013; Roy and Guntuku, 2016; Gao et al., 2017; Lee and Abu-El-Haija, 2017; Chen et al., 2018, 2021; Li et al., 2019; Liu et al., 2019a; Ma et al., 2019; Du et al., 2022; Yi et al., 2022
Sports	Textual	Sanchez et al., 2012
	Visual	Ramezani and Yaghmaee, 2016

Publications may appear multiple times if more than one feature modality is used.

of recommendations and offers the potential to explain why a video is suggested.

Restricted Boltzmann Machines (RBM) are a type of *neural network (NN)* used in Hazrati and Elahi (2021) to learn the latent representation of videos in a feature space. Visual features are employed for model training, capturing complex connections in the input features. The model assigns different weights to individual input features, reflecting their representativeness of the video content.

An alternative approach for content representation is to classify videos by topic using extracted features as input. For instance, in Luo et al. (2008), multi-modal features are synchronized to learn topic representations for news videos, while Zhu et al. (2013) introduces a topic-modeling approach for movies.

A special task of VRS is the recommendation of *micro videos* (sometimes *short videos*), commonly found on social network platforms, like TikTok.[11] These videos have a small duration (usually seconds to minutes) and limited textual descriptions, requiring systems to rely on automatically extracted features for their recommendations.

Multi-Modal Graph Contrastive Learning (MMGCL) is introduced in Yi et al. (2022) to learn multi-modal representations for micro videos. This self-supervised method employs augmentation techniques and negative sampling to achieve accurate representations, considering the correlation between different modalities. Similarly, in Du et al. (2022), the modality correlation is explored using a *Cross-modal Graph Neural Network* to encode and aggregate cross-model information, enabling the creation of modality-aware representations for users and micro videos. The self-supervised learning approach used is *Cross-modal Mutual Information Fusion*, which captures the correlation between video modalities.

The *VideoReach* system (Mei et al., 2007, 2011), addresses the integration of multi-modal features for video representation. It combines manually crafted and automatically extracted aural, textual, and visual features, mapping them to textual descriptions for compatibility with textual recommendation methods. The system assigns predetermined weights to feature types, focusing more on textual features due to their rich information content. These weights are individually adjusted based on user feedback, measured through the *Click-Through-Rate (CTR)* that captures user interactions like selecting, pausing, or seeking videos. This feedback helps adapt modalities' relevance and results in improved video representations.

3.2. Content-based recommenders

Content-based recommenders, also known as *Content-based Filtering (CBF)*, utilize item characteristics or features that users are interested in to find unseen items with similar attributes and present those as recommendations (Nikolakopoulos et al., 2022). The aim of CBF is to leverage the commonalities of item features that have been relevant to a *target user*, i.e., a user for whom a recommendation is computed, in the past, by suggesting items with high overlap in terms of similarity, determined by various similarity functions (Adomavicius and Tuzhilin, 2005).

Analyzing the publications on video recommenders revealed that content-based recommendations are predominantly computed using *supervised*, *unsupervised*, and *self-supervised* learning

11 https://www.tiktok.com

TABLE 3 Content-based VRS approaches classified by applied technique and algorithms.

Type	References
Supervised learning	Luo et al., 2008; Zhu et al., 2013; Chantanurak et al., 2016; Elahi et al., 2017, 2020, 2021; Lee and Abu-El-Haija, 2017; Deldjoo et al., 2018a; Tavakoli et al., 2020; Hazrati and Elahi, 2021; Rimaz et al., 2021; Leite et al., 2022
Unsupervised learning	Wu et al., 2008; Davidson et al., 2010; Sanchez et al., 2012; Niu et al., 2013; Deldjoo et al., 2016, 2018b; Ramezani and Yaghmaee, 2016; Lu et al., 2017
Self-supervised learning	Mei et al., 2007, 2011; Covington et al., 2016; Gomez-Uribe and Hunt, 2016; Chen et al., 2018; Kaklauskas et al., 2018; Li et al., 2019; Jiang et al., 2020; Chakder et al., 2022; Du et al., 2022; Gong et al., 2022; Pingali et al., 2022; Yi et al., 2022; Mondal et al., 2023; Xiao et al., 2023

approaches. Table 3 classifies publication by these approaches. While supervised approaches determine whether an item is relevant or irrelevant to the target user, unsupervised approaches seek the most similar content based on the distance to a *seed* in the embedding space, where the seed describes the current user preference. Self-supervised techniques predominantly involve *Deep Learning* models to learn content structures for predicting item relevance.

In the following, the publications and technical approaches to computing content-based recommendations are discussed in detail.

3.2.1. Supervised learning

Supervised learning algorithms for content-based recommendation take the feature descriptions of items and user preferences (often defined as *user profiles*) as input to predict whether an item is relevant with respect to individual preferences. It comprises *classification*, i.e., the assignment of items to predefined categories like relevant/irrelevant, and *regression* analysis, i.e., the prediction of a numerical value like a user rating. Thereby, different features, feature representations, encoding of user preferences, and classification techniques are applied, depending on the context.

A predominantly used algorithm in content-based video recommendation is *k-Nearest-Neighbors (kNN)* (Luo et al., 2008; Zhu et al., 2013; Elahi et al., 2017, 2020, 2021; Lee and Abu-El-Haija, 2017; Deldjoo et al., 2018a; Hazrati and Elahi, 2021), which identifies the *k* most similar items, given a distance metric applied to the item features (Jannach et al., 2011b). Items are more similar, the lower the distance between them. In Chantanurak et al. (2016), this approach was used to recommend learning videos from YOUTUBE. It uses keywords from course metadata in a *Learning Management System (LMS)* as search queries to obtain a video selection and the available video keywords. Those are transformed to a TF-IDF representation, used for the kNN recommendation.

Besides comparing the similarity between video items, often a user profile reflecting the individual user preferences is used to identify similar videos. Mostly, this profile is based on past video consumption and represented in the same embedding space as the videos, which enables computation of the distance between them. An elaborated example is presented in Zhu et al. (2013), using a two-tower approach for the recommendation. In the video representation stage, a topic model based on textual and visual features is learned to describe the video. In the second stage, the user is described as a topic model based on their watch history. Relevant videos are identified by the minimal distance between the user model and topic models of videos.

Another supervised approach for content-based video recommendation is *Random Forest* (Ho, 1995). This machine-learning approach combines multiple decision trees to classify an item as relevant or irrelevant. The final decision is made through a majority vote. An example is presented in Tavakoli et al. (2020), where a model determining the relevance of learning videos to a user based on their current knowledge level and job skill requirements is developed, aiming to assess if a video matches a skill description in the educational video recommender.

3.2.2. Unsupervised learning

Unsupervised learning algorithms for content-based recommenders extract patterns and relationships from unlabeled data to provide meaningful insights and recommendations without predefined categories. Clustering is one approach, which groups items such that items assigned to the same group (cluster) are more similar compared to others. For content-based recommendation, this approach is used to identify similar items to a seed or user preferences represented in the same embedding space. Any kind of content representation can be taken into account (see Section 3.1), and the approach is applicable to a variety of domains, e.g., for clustering sports videos based on recognized human actions (Ramezani and Yaghmaee, 2016) or using the identified topic of videos (Wu et al., 2008).

A popular clustering approach for video recommendation is *k-Means* (Wu et al., 2008; Deldjoo et al., 2016, 2018b; Ramezani and Yaghmaee, 2016), which is an iterative algorithm that assigns items to one of *k* clusters, such that the distance between the centroid (cluster center) and the item is minimized, given a distance metric (Jannach et al., 2011a). For a standard recommendation approach, clustering can be involved to identify the most similar cluster based on a user's context and recommend videos from that cluster that the user has not seen yet. Furthermore, clustering can also be beneficial in a two-stage recommendation process, where it helps generate an initial set of candidates from a large video catalog (Davidson et al., 2010). By using a fast clustering algorithm, the overall performance can be improved by prefiltering the videos, which are then ranked using a more accurate but slower algorithm. By taking neighboring clusters into account, the exploration of additional topics is favored, which can further improve the user experience (Wu et al., 2008).

Users may have distinct individual reasons for being interested in a video. For instance, one user appreciates the plot, while others are interested in the actors. In Lu et al. (2017), these factors are considered. Videos are clustered using a multinomial vector representation, where different topics are assigned to the same video with corresponding weights. Users are also modeled in this space based on their watch history, enabling the identification of the nearest cluster and recommending videos from that cluster.

Spectral clustering is an algorithm from the graph theory using eigenvalues of a similarity matrix to group items (Ng et al., 2001). In Niu et al. (2013), it is used to recommend videos based on the user's mood. The videos in this approach are clustered by their *affective* properties (see Section 3.6.1).

Another unsupervised approach is the usage of a *Hidden Markov Model (HMM)* (Baum and Petrie, 1966). In Sanchez et al. (2012), this has been used to recommend Olympic Games transmissions given a user profile and manually created video annotations. The system builds a user profile modeling user interests with weighted factors for preferences such as preferred sports and athletes. The profile evolves continuously based on consumed content using an HMM capturing the interest in specific videos as *hidden states*. The HMM parameters are used in a Bayesian inference step, to calculate the probability of video relevance to the user.

Also, *association rule mining* (Liu et al., 1998), which is a data mining technique that discovers relationships and patterns within large datasets based on item co-occurrence, can be used for content-based video recommendation. In Davidson et al. (2010), the approach is used to calculate a *relatedness score* of other videos in the catalog given a video watched by the user. This score represents the relations between videos as a directed weighted graph. A candidate set of items is then generated considering a limited transitive closure within a specified distance. The candidates are subsequently ranked based on various properties such as video quality (e.g., recentness and general popularity), user specificity (compatibility with the user watch history), and diversification (removal of similar videos to promote serendipity).

3.2.3. Self-supervised learning

Self-supervised learning algorithms for content-based recommenders use automatically generated item embeddings (Chen et al., 2022b,c) as input to predict recommendations without requiring explicit user-item interactions. Those systems apply different types of neural networks to predict user ratings for videos using a variety of inputs.

In Kaklauskas et al. (2018), personal user characteristics are combined with real estate advertising videos in a *neuro decision matrix*, which is a cognitive framework employing neural network models to analyze complex patterns and data inputs enabling personalized decision-making. It is used to deliver personalized video clips showcasing properties matching individual preferences.

Pooling the video embeddings of positively rated videos using the feature-wise mean to obtain a user embedding is applied in Pingali et al. (2022). These user embeddings and embeddings of unseen videos are fed into a *Siamese neural network*, which is a neural network capable of comparing the similarity between two patterns. By utilizing a regression function, the method predicts ratings for similar videos. Using a *Graph Attention Network* (Chakder et al., 2022) or *Graph Convolutional Neural Network* (Mondal et al., 2023) to develop the regression system and extend movie embeddings with further latent features, the accuracy of this approach can be improved.

In Chen et al. (2018), a deep network-based method for the prediction of user clicks on micro videos is presented. The *Temporal Hierarchical Attention at Category- and Item-Level (THACIL)* network uses a combination of temporal windows to capture short-term dynamics of user interest, and multi-grained attention mechanisms to describe the diverse user interest. While category-level attention describes the diverse interest of users, fine-grained user interests are described with item-level attention. Using a hierarchical attention mechanism, short-term and long-term properties of user behavior are modeled.

Micro video recommendation faces the challenge of dynamic and diverse user interests, leading to the development of various solutions. One baseline strategy uses time decay to reduce the significance of videos watched further back in the past. An advanced version employs a temporal graph-guided network, as described in Li et al. (2019), to predict the click probability of videos. This model combines past user behavior with diverse topic preferences, considering both engaging and uninteresting videos from the user's viewing history. Furthermore, the model incorporates the notion of varying interest levels in topics, where actions such as liking a video are given higher importance than merely watching it.

Using a static time decay heuristic fails to consider personalized and individual preferences, where older videos might be more important for some users. In Jiang et al. (2020), a *Multi-scale Time-aware user Interest modeling Network (MTIN)* is proposed to address this issue. MTIN incorporates a parallel temporal mask network to capture varying importance over time. Additionally, the model utilizes a grouping approach for videos and assigns users to multiple interest groups, allowing for a more accurate representation of their diverse preferences.

To handle the dynamically changing user interests in micro-video applications, a real-time re-ranking solution was proposed in Gong et al. (2022). Recognizing that traditional server-side models might not capture short-term preferences from user interactions with minimal delay, the approach suggests deploying a lightweight edge-side model on the client side to re-rank the recommendations after each user interaction. This approach divides roles, utilizing server-side models for complex, enduring preferences, and enabling client-side models to incorporate immediate feedback for real-time adjustments.

CTR prediction, i.e., the anticipation of the following user action, is a challenge in video recommendation (Liu et al., 2020). In this context, the goal is to foresee a user's upcoming video choice based on their past interactions. Deep learning models based on the *Embedding and Multilayer Perceptron (MLP)* paradigm are commonly used for this task. These models map input features to low-dimensional embedding vectors, which are then transformed and concatenated in MLP layers to capture non-linear relationships among the features (Zhou et al., 2018). Nonetheless, this approach struggles with diverse user interests. For instance, if a user watches action, romantic, and science-fiction movies, merging all genres into a single representation might overlook genre-specific relevance due to the user's varied history.

To address this, the concept of *Deep Interest Networks (DIN)* was introduced in Zhou et al. (2018). DIN acknowledges that a portion of a user's interests can impact their subsequent actions, like choosing a movie. It dynamically computes the interest by considering historically significant actions related to a candidate item. A local activation unit with soft search identifies relevant portions of user history. A weighted sum pooling method generates an interest representation for the candidate item, assigning greater weights to more relevant segments. To incorporate user feedback into predictions using DIN, the *Preference Matching Network (PMN)* model was presented in Liu et al. (2020), following the idea that users are more inclined to accept candidate items that resemble videos they have positively rated. PMN first calculates similarity weights between a candidate video and the user's interaction history. Then, a weighted sum pooling of the user's feedback is calculated to determine their preference for the given candidate.

The exploration of user interest for CTR prediction as an extension to relying exclusively on historical behavior was suggested in Chen et al. (2022a). By explicitly modeling item relations and including them in the network for embedding user interest, recommendation quality can be improved.

In Xiao et al. (2023), a solution to tackle the cold start problem for new users was presented. The solution incorporates information from similar users in the social network. If the video platform shares users with a social network, a social graph can be created, capturing relationships such as friendships or common interest groups. Through clustering, similar groups of users can be identified. By aggregating the interests of these social groups with user features, the accuracy of personalized recommendations can be enhanced.

3.3. Collaborative filtering

Collaborative Filtering (CF) is based on the concept that users with similar preferences in the past will continue to have similar preferences. Hence, CF exploits past ratings to suggest unseen items by considering items liked by users with similar preferences (Ricci et al., 2015). The core assumption is that similar users share interests in similar items, and analogous items are favored by similar users (Nikolakopoulos et al., 2022). This involves identifying similar users, often termed as *neighbors*, by calculating the similarity of past ratings using measures like *Pearson correlation*, *cosine similarity*, or *Spearman's rank correlation coefficient* (Jannach et al., 2011a). Ratings can be explicit (direct user ratings or subscriptions) or implicit (derived from user behavior like viewing time) (Davidson et al., 2010; Koren et al., 2022).

For video recommendation, CF provides an intuitive approach, recommending unseen videos based on the preferences of users with similar interests. Table 4 groups various systems using this approach by their techniques. The summary shows that similar to the content-based recommendation (see Table 3), supervised, unsupervised, and self-supervised learning methods are widely used to compute recommendations.

In the following, the publications and algorithmic approaches of applying CF for video recommendation are discussed in detail.

TABLE 4 Collaborative filtering VRS approaches classified by applied algorithms.

Type	References
Supervised learning	Arapakis et al., 2009; Dias et al., 2013; Choi et al., 2016; Okubo and Tamura, 2019
Unsupervised learning	Wang et al., 2014; Ferracani et al., 2015; Katarya and Verma, 2016; Katarya, 2018; Tohidi and Dadkhah, 2020
Self-supervised learning	Hongliang and Xiaona, 2015; He et al., 2017; Rybakov et al., 2018; Yan et al., 2019
Further approaches	Baluja et al., 2008; Koren et al., 2009; Chen et al., 2015, 2019

3.3.1. Supervised learning

Supervised learning in content-based and collaborative filtering diverges mainly in their used input. While CBF employs item content features to find similar items, CF operates on a user-item rating matrix along with the target user. CF utilizes nearest neighbors algorithms on the matrix to identify users who are similar to the target user. The process typically involves three steps (Dias et al., 2013): (1) Similarities between the target user and others are computed using ratings and a similarity metric. (2) The most similar users, known as neighbors, are selected. (3) Item ratings are predicted from the weighted average of neighbor ratings. While explicit ratings for the video or segments of a video (Dias et al., 2013) are frequently used, recommendations can as well be based on implicit ratings, for example, by applying emotion recognition to derive user preferences (Arapakis et al., 2009; Choi et al., 2016; Okubo and Tamura, 2019).

3.3.2. Unsupervised learning

Video-based collaborative filtering often starts with clustering to decrease the search space of the model-based approach. Optimization methods are then used on similar user clusters, rather than the entire user space, to enhance scalability. Given the target user, the nearest cluster is identified, and video ratings are predicted using a weighted average of other users in the cluster.

Many methods use the k-Means algorithm for clustering similar users and enhancing the accuracy with varied optimization techniques. For instance, in Katarya and Verma (2016) *Particle Swarm Optimization (PSO)* is applied for improved cluster centroid assignment. The *Artificial Bee Colony (ABC)* algorithm optimizes user-cluster assignments (Katarya, 2018). In Wang et al. (2014), k-Means is paired with genetic algorithms in a two-step approach. Firstly, *Principal Component Analysis (PCA)* condenses data dimensions by removing less significant data. Secondly, this dense data is clustered to identify similar users.

Furthermore, the clustering itself can be improved. In Tohidi and Dadkhah (2020) evolutionary algorithms based on k-Means were used for this purpose. Alternatively, the *Fuzzy C-means (FCM)* algorithm, permits users to belong to multiple clusters with varying degrees of membership (Ferracani et al., 2015; Katarya, 2018). FCM optimally assigns users to these clusters, promoting a diverse user profile representation.

3.3.3. Self-supervised learning

Self-supervised learning in collaborative filtering generates user vector representations reflecting their interests. Embeddings of users are compared using a distance metric to find target user neighbors. The weighted average of the neighbor's ratings is used to predict the item ratings used as recommendations.

In Hongliang and Xiaona (2015), a *Deep Belief Network (DBN)* quickly extracts user features, e.g., preferred genres and movie ages. User ratings are encoded as a binary matrix, where each movie corresponds to a column, and each rating value option is represented by a row (1 for rated, 0 for unrated). This matrix is then used as input for the DBN to generate a user feature vector. The feature vectors for all users are used to find nearest neighbors using the *Euclidean distance*.

Without explicit ratings, user preferences can be inferred from interactions as implicit feedback. The *Neural network-based Collaborative Filtering (NCF)* presented in He et al. (2017), takes the user and item ids as input features, converting them to binarized sparse vector with one-hot encoding. In the embedding layer, the item vector is projected to a dense representation, which is then fed into the multi-layer network for a prediction score. This score, obtained from the final layer, gauges video relevance for the target user.

In Rybakov et al. (2018), a two-layer neural network is trained to predict users' upcoming video selection. The model is designed to forecast videos to be consumed within a specific time frame, such as the upcoming week, leveraging the insight that predicting the next item is more accurate than random future items (Covington et al., 2016). This approach effectively captures both short-term trends, such as current events like the COVID-19 pandemic, and long-term user preferences. The model combines a *predictor* for currently popular items and an *auto-encoder* for static user preferences in a feed-forward neural network. The system is retrained daily to adapt to changes. The recommendation precision is improved by considering consumption dates through time decay, approximated through a convolutional layer.

As sparse user ratings can negatively impact the recommendation quality, the usage of sentiment analysis on free-text reviews is suggested in Mahadevan and Arock (2017) to address this issue. NLP techniques are used to deduce numerical ratings from credible reviews, which are then used in the recommendation process. Experiments showed improvements compared to the direct usage of ratings from the datasets. This highlights the potential of mapping text reviews to ratings for more meaningful user interest understanding than numeric ratings alone.

In video recommenders, personalized suggestions are typically based on user data like viewing history. However, in cold-start situations, where data is scarce, sharing information with other platforms or social networks can enhance user profiles. In Deng et al. (2013), two strategies were evaluated: (1) directly incorporating user profiles from an auxiliary platform to enrich the target platform, and (2) transferring user relationships (i.e., behavioral similarity) from the auxiliary to the target platform. This information was combined with user interactions on the video platform to compute personalized recommendations. Experiments revealed certain aspects of auxiliary profiles, such as shared articles and registration info, were more valuable than others. While integrating all data did not always improve accuracy and sometimes performed worse than relying solely on the target platform's sparse profile, selectively integrating relevant information from the auxiliary platform showed potential for performance improvement.

The discrepancy of user interests in different services, stating that user interest features include cross-site commonalities and site peculiarities, is observed in Yan et al. (2019). The study revealed, that *multi-homed users*, i.e., users using multiple services, have inconsistent and independent preferences in different services. Analogously, *multi-homed videos*, i.e., videos uploaded to multiple services, enable sharing of user interests across services. To tackle this, the study employs the *Deep Attentive Probabilistic Factorization (DeepAPF)* model, which splits user embeddings into common and site-specific parts, adapting feature weights via an attention mechanism. This approach captures both shared and unique user preferences across services.

In the domain of e-learning, cross-correlation of videos can be applied to leverage the use of videos across different courses, emphasizing the correlation of knowledge between courses (Zhu et al., 2018). This is achieved through a two-step approach: (1) CF is used to form a seed set of pertinent videos based on learner interactions like video view duration and navigation. (2) The degree of relevance between videos is computed using a cross-curriculum knowledge map, and a random walk algorithm is employed to measure the degree of relevance. This generates video subgraphs that contain video recommendations aligned with both learner preferences and the knowledge relevance of the video content.

3.3.4. Further approaches

Adsorption is a graph-based semi-supervised learning approach that leverages user-video preferences for video recommendation (Baluja et al., 2008). It propagates known user preferences (labeled nodes) to unknown preferences (unlabeled nodes) based on the view history of users. Users and videos are represented as nodes in the graph, which are linked if users viewed them. Videos for recommendation are determined by identifying videos connected by short paths through other users.

Singular Value Decomposition (SVD++), forms a powerful method for collaborative filtering that improves traditional matrix factorization (Koren et al., 2009). It includes implicit feedback and explicit user/item biases. The technique factors the user-item rating matrix into lower-dimensional matrices representing latent factors. These factors capture underlying features. The model approximates the original ratings by multiplying these matrices. To consider implicit feedback, a weighted regularization term is introduced, which considers the confidence of observed user-item interactions. This prioritizes highly relevant data. Explicit user/item biases handle inherent rating data biases, capturing individual user tendencies and item popularity.

In Chen et al. (2015), an *Artificial Immune System (AIS)* for CF is introduced. AIS mimics biological immune systems, comprising *antigens* (unclassified training data) and *antibodies* (generated in response to antigens). These antibodies construct specialized *immune networks* signifying their similarity to antigens, representing specific training data. After training, the final immune network predicts user ratings for a target user (antigen). This involves identifying nearest neighbors via similarity assessment of user groups (immune networks) and users within those groups (antibodies). By leveraging this immune system-inspired approach, accurate predictions can be made for the target user's ratings.

TABLE 5 Hybrid VRS approaches classified by applied algorithms.

Type	References
Matrix factorization	Cui et al., 2014; Roy and Guntuku, 2016; Kvifte et al., 2021; Wang et al., 2021
Deep neural networks	Wang et al., 2015; Gao et al., 2017; Wei et al., 2017; Liu et al., 2019a; Chen et al., 2021
Multi-task learning	Ma et al., 2019; Zhao et al., 2019; Tang et al., 2020; Zhuo et al., 2021; Song et al., 2023
Further approaches	Öztürk and Kesim Cicekli, 2011; Vizine Pereira and Hruschka, 2015; Abbas et al., 2017; Liu et al., 2019b; Kim et al., 2021

To handle the problem of unavailable explicit ratings, *Interest Preferences of Categories (IPoC)* can be deduced as implicit ratings from user logs (Chen et al., 2019). View times of short videos are used to determine ratings, reflecting user interest in specific categories through weighted video consumption times. These ratings are then used to fill a rating matrix for CF using matrix factorization. By weighing values higher for frequently consumed categories and factoring IPoC confidence, rating accuracy is enhanced.

3.4. Hybrid recommenders

Hybrid recommendation approaches combine various strategies to overcome the limitations of single recommendation strategies (Nikolakopoulos et al., 2022). Various hybridization designs are commonly employed (Jannach et al., 2011c). Firstly, the *parallel* design involves implementing multiple systems independently and combining their recommendations. Secondly, the *pipelined* design merges different approaches by using the output of one system as input for the subsequent recommender. Lastly, the *monolithic* design integrates diverse input data, e.g., item features and user ratings, into a single model.

The fundamental principle of hybrid recommenders is the integration of multiple strategies, like content-based and collaborative filtering, to overcome the limitations of individual methods, and enhance the accuracy and diversity of video recommendations. Hybrid systems commonly tackle data sparsity, scalability, and cold-start problems. An overview of the technical approaches used in publications is shown in Table 5.

In the following, the publications and algorithmic approaches for hybrid video recommendations are discussed in detail.

3.4.1. Matrix factorization

Matrix factorization is an embedding model used to predict user ratings for unrated items. A characteristic of matrix factorization is the transformation of users and items in the same vector space, where both are clustered based on the similarity of latent factors (hidden features).

One option is to represent social media users and videos in a common attribute space (Cui et al., 2014). This method involves enriching videos with social aspects, like demographic data of viewers, and user profiles with content information from watched and liked videos. Experiments detected the appropriate balance of content and social attributes, favoring social attributes. This monolithic design aligns users and videos in a single attribute space, focusing on similarity-based matches for recommendations. For sparse videos, content similarities share social attributes, and user relationships share content attributes. The design effectively handles cold start for both items and users by mapping them to videos with similar content and common user relationships.

The model described in Roy and Guntuku (2016) emphasizes users' emotional influences on video preferences. It enriches collaborative data with recognized emotions users experience while watching videos. By integrating emotions, the model gains latent factors capturing emotional user-video connections. These latent factors are then used in a factorization method for rating predictions.

To improve the accuracy of recommendations in the presence of cold start and sparse ratings different approaches were suggested. In Kvifte et al. (2021), the usage of aggregated content data (visual features and word frequency in subtitles) and user ratings to predict recommendations via matrix factorization was presented. In Wang et al. (2021), a two-tower model is proposed to improve cold starts. One tower learns user embeddings from watch history, while the second tower learns item representations from metadata (e.g., genres, actors, and synopsis) and movie cover art. An attention layer weighs features based on item importance. Matrix factorization approximates user preferences with embeddings.

3.4.2. Deep neural networks

Hybrid video recommenders using deep neural networks often aim to enhance recommendation accuracy by incorporating content features and user ratings. *Collaborative Deep Learning (CDL)* unites deep representation learning for content and collaborative filtering for ratings (Wang et al., 2015). This allows for a two-way interaction between the input information. Content features improve CF predictions and video ratings support feature learning using a *stacked denoising autoencoder (SDAE)*, which is a deep learning model that learns a hierarchical representation of data by removing noise and reconstructing clean input. Using this model, CDL generates accurate rating predictions for user-video pairs.

In Wei et al. (2017), the cold start problem is tackled by integrating an SDAE into the CF model *timeSVD++*. This model considers user preferences, item features, and temporal rating dynamics. The process starts by extracting and processing movie plots for relevant words. A bag-of-words vector captures item similarity. These vectors train the SDAE to extract item content features. The trained features are the input for the CF model that predicts the ratings of items with few or no ratings based on similar items which are already sufficiently rated.

Dynamic Recurrent Neural Networks (DRNN) (Gao et al., 2017) fuse dynamic user interest with content details. The system merges video semantics (textual and visual description), user interest from history, and user relevance (collaborative aspect) for similar user discovery. It adapts for single or cross-network use, possibly incorporating social networks for improved accuracy. Videos are represented in a semantic space using multi-modal features, and a common interest space connects semantics and user interest. An RNN models dynamic user interest over time, using a ranking loss constraint in the final RNN state to consider user relevance. This model acts as an interest network, harmonizing these sources to understand dynamic user preferences and provide interpretable user-video recommendations.

Hybrid approaches have also been implemented for micro-video recommendation. In Liu et al. (2019a), a model predicting if users will finish and like a video subsequently is described. The prediction model is learned from user interaction and multi-modal item feature data. To enhance the accuracy of predictions, an ensemble method is employed, utilizing individually predicted ranks from multiple prediction models. Notably, each model takes into account different time frames of the user's interaction history, leading to a more comprehensive understanding of user preferences and behavior.

In Chen et al. (2021), a method to combine various user interest representations for micro-videos and movies is presented. This approach fuses different representations of user interest, including the overall user profile, item and category-level representations, and collaborative data using a DNN. The outcome is a unified representation synthesized from different preference sources.

3.4.3. Multi-task learning

Multitask learning (MTL) is a machine learning approach that trains one model for multiple related tasks, boosting performance through shared representations (Tang et al., 2020). In video recommenders, objectives can be diverse and sometimes conflicting. In that sense, the same system can have engagement objectives like clicks and watch time, while also considering user satisfaction indicated by likes or ratings (Zhao et al., 2019). MTL can help to tackle this challenge.

A model for combining three optimization goals, namely the partial order between videos, CTR, and prediction of the sequentially clicked video, was presented in Zhuo et al. (2021). Using a behavior-aware graph convolution network, the system differentiates user behaviors to reflect the influence between users and videos. Behaviors (e.g., clicks, watch duration, and ratings) are mapped to scores, adjusting interaction weight based on strength, where higher scores resemble greater user interest. Those weightings are merged into the embedding space of users and items. The model objective of learning is to estimate the probability of the target user choosing each of the available videos.

In Zhao et al. (2019), the ranking phase of video recommendation was enhanced by incorporating the *Multi-gate Mixture-of-Experts (MMoE)* architecture for MTL. MMoE has a shared bottom layer and separate expert layers per objective. The expert layers learn task-specific data from inputs. Gating layers for each task incorporate expert and shared input. The expert layer output is fed into a task layer predicting binary objectives (e.g., clicks and likes) or regression tasks (e.g., watch time, and ratings). In Song et al. (2023), MMoE is adapted for playback prediction, based on user history, embeddings, and playback time.

Those systems might suffer from the implicit *selection bias*, where the interaction logs used for model training do not capture whether users clicked on a recommended video because it genuinely matched their preferences or because it was simply ranked higher, potentially causing more relevant videos in the catalog to be overlooked. To mitigate this bias, a *shallow tower* alongside MMoE was added in Zhao et al. (2019). This tower uses inputs contributing to the selection bias (e.g., video position and device data) and integrates its output into the main model's final logit. This reduces bias and improves fairness and system efficacy.

Progressive Layered Extraction (PLE), presented in Tang et al. (2020), forms an MTL approach improving shared learning efficiency while reducing *negative transfer* and the *seesaw phenomenon*. Negative transfer in RS occurs when unrelated objectives lower performance compared to single-task systems. The seesaw phenomenon is the trade-off between improved performance for one task and a decline in others in MTL. PLE is built on the *Customized Gate Control (CGC)* model, segregating shared and task-specific experts to avoid parameter interference. Task-specific experts focus on learning distinct knowledge, receiving input from their expert network and the shared expert network through a gating network for dynamic fusion. PLE extends CGC to a generalized model with multi-level gating networks and progressive separation routing, stacking CGC expert networks and creating extraction networks. Each extraction network receives fused outputs from lower-level networks, gradually learning deeper semantic representations and extracting higher-level shared information. By separating task parameters in upper layers, PLE enables the extraction of deeper semantic representations for each task, fostering generalization.

3.4.4. Further approaches

A combination of the CF graph algorithm *Adsorption* with content-based similarity to improve the quality of recommendation was presented in Öztürk and Kesim Cicekli (2011). The system constructs a user-item graph, with users and items as nodes and weighted edges indicating interactions (e.g., likes). Items are initially labeled as relevant or unknown for each user. Adsorption spreads labels from labeled items to nearby ones, indicating relevance. Unrated videos reached via the graph are recommended. To improve the recommendations, the CF results are refined by including videos with similar content features, replacing less relevant suggestions.

Combining CF with *Demographic Filtering (DF)* (user profile creation from demographic characteristics) offers one possibility to address the cold start problem (Vizine Pereira and Hruschka, 2015). The *Simultaneous Co-Clustering and Learning (SCOAL)* algorithm uses video and user characteristics to create prediction models for different co-clusters, aiding users with minimal ratings by assigning them to the closest cluster. For users without any ratings, the cluster description and demographics determine the best prediction models. The first approach estimates the probability distribution for each co-cluster and calculates the predicted rating as a weighted sum, while the second, more resource-intensive method, constructs a video-by-video classifier involving only users who have rated the video.

The problem of sparse user ratings is addressed in Liu et al. (2019b) by computing user-video similarities using collaborative user similarity from ratings and content representation, which includes genre similarity and word embeddings from textual descriptions. These two similarities are fused using an adjusted weighted sum, which considers varying rating data importance. Ultimately, kNN recommends most similar videos based on these fused similarities.

3.5. Group recommenders

Group recommendation involves recommending items to a collective group rather than individual users, assuming the preferences of group members are known or can be obtained through recommender systems (Felfernig et al., 2018; Masthoff and Delić, 2022). Aggregating individual user models becomes a challenge in this approach, adding complexity to the recommendation process. An example of group recommendation is recommending a TV program that satisfies all viewers in a family watching TV together (De Pessemier et al., 2016).

In group video recommendations, the aim is to unite diverse individual user models with different strategies (Masthoff and Delić, 2022). For instance, in interactive television, the selection of programs should take into account the satisfaction of the entire group, not just the preferences of a single individual. Group recommenders face the particular challenge of balancing individual member satisfaction while suggesting items that align with the overall group preferences.

The *PolyLens* system(O'Connor et al., 2001), an extension of *MovieLens* (Harper and Konstan, 2015), focused on group movie recommendations. Users could create groups and receive movie suggestions based on collective group preferences rather than individual ones. Guided by a social value function, the process aimed to maximize the overall happiness of the group, gauged as the minimum happiness score among members. Recommendations excluded movies already viewed by some group members. Group suggestions were created by merging individual users' recommendation lists and ranking them based on least misery or decreasing social value. This method proved effective for smaller groups (2-4 people) with participants perceiving the generated recommendations as valuable and agreeing on their usefulness.

As an alternative to merging recommendation lists, the aggregation of user profiles to generate recommendations was presented in Yu et al. (2006). This technique is geared toward suggesting TV programs for groups watching TV together. The merging process combines vectorized feature descriptions of all group members' profiles by minimizing the total distance between them, aiming to retain the most common characteristics. To adjust for individual preferences, weight normalization is applied to the merged profile vector. By merging profiles and considering the collective characteristics, the system creates tailored recommendations for an enhanced TV experience.

The recommendation of movies for on-demand cinemas presents a unique application of context-aware group recommendation systems (Xue et al., 2019). This application focuses on combining classic cinemas with on-demand streaming, allowing groups to select movies in cinema rooms with specific equipment. Recommendations are essential for aiding guest decisions, though personalization is challenging due to the unknown and anonymous audience. The system addresses this by leveraging contextualization, considering temporal and spatial characteristics. Attendees are assumed to be local, and movie preferences vary based on the temporal aspect. Each cinema is expected to have its unique characteristics influenced by its environment captured by *Points of Interest (POI)* nearby. By collecting cinema activities like selected movies, time, and location, individual cinema profiles are created, integrating POIs, movie details, and ratings. Using this data, the system employs CF to model temporal and spatial dynamics. Temporal dynamics cover the *Periodic Effect* (common viewing patterns by time, day, and season), *Recency Effect* (preference for new movies), and *Audience Crowd Drifting Effect* (varying composition of audiences by time, such as couples or families). The spatial context is modeled through the Spatial Neighboring Effect (similar audiences in cinemas with similar POI patterns) and the *Spatial Popularity Effect* (differing regional movie popularity). This enables the prediction of movie ratings for specific cinemas at given times.

3.6. Further aspects

This chapter delves into various aspects of video recommenders, including the incorporation of affective signals like unconscious expressions and body language of users into RS, video recommendations tailored to consumption contexts, scenarios involving only certain parts of longer videos, publicly available datasets for VRS development, and an overview of metrics used to evaluate the recommendation quality.

3.6.1. Affective computing

Affective computing aims to integrate human-like capabilities of perceiving, interpreting, and generating affect features, like emotions and mood in computers (Tao and Tan, 2005). This involves using sensors that capture diverse aspects of human behavior, such as gestures, voice, and heart rate, allowing computers to understand and respond in a friendly and intelligent manner. In recommender systems, this data enhances user profiles and feedback with unique information.

Using affective sensory data to automatically retrieve feedback is a popular method for determining user preferences in various video domains, such as TV program recommendation (De Pessemier et al., 2016), movies (Okubo and Tamura, 2019; Bandara et al., 2021), and advertisements (Choi et al., 2016; Kaklauskas et al., 2018; Kim et al., 2021). Facial expressions of users captured with webcams while watching videos provide more expressive opinions compared to simpler approaches, such as assuming that watching a video indicates liking (Arapakis et al., 2009; Choi et al., 2016; De Pessemier et al., 2016; Kaklauskas et al., 2018; Okubo and Tamura, 2019; Kim et al., 2021). Studies have shown positive correlations between identified smiles of users and video appreciation (Arapakis et al., 2009; Okubo and Tamura, 2019), but the correlation between emotions and ratings remains inconclusive in some cases (Diaz et al., 2018). Using DNNs, the emotion of users can be detected instantly to identify dynamic preferences and decide if recommended videos are appropriate (Choi et al., 2016; Kim et al., 2021). Since those approaches do not rely on a user history or a pre-existing profile, they offer a solution for cold-start situations in which the user is unknown.

In Kaklauskas et al. (2018), an affective VRS is designed to aid a variety of potential real estate buyers in discovering suitable properties. The system presents personalized property videos to users and records their facial expressions during viewing to gauge their emotional response. This data is utilized to determine whether to play another video clip and to identify the most suitable video from the catalog for the user.

Several VRS incorporate affective data for recommendations. In Roy and Guntuku (2016), the emotional connection between users and videos is modeled, suggesting users prefer videos they can emotionally connect with. To forecast emotional user reactions, a multi-label *Support Vector Machine (SVM)* classifier is used. SVM is a supervised machine learning method that determines an optimal decision boundary to classify data into classes, maximizing the margin between the closest data points of each class.

A related idea is applied in Niu et al. (2013) to recommend videos based on the user's current mood. The system utilizes a valence-arousal graph to autonomously learn affective attributes from videos. Valence signifies emotions from "pleasant" to "unpleasant," while arousal measures the intensity of emotions from "excited" to "calm," on a continuous scale. Recognizing that users' moods are dynamic and not static, the system captures users' affective traits within a session, encompassing sequentially watched videos. This approach assumes that the emotional impact of previously viewed videos influences the selection of the next video.

The usage of *Electroencephalograms (EEG)*, which measure brain neural activity, to capture user emotions and attention while watching videos is explored in Bandara et al. (2021). Using headbands, the brain activity of test users watching movie trailers was recorded. The EEG signals were classified into various emotional states, considering engagement and attention levels. Through EEG analysis, the system predicts video clip relevance to users based on their emotional and attention responses, which are then used for generating video recommendations.

In Leite et al. (2022), an affective virtual learning environment for algebra is examined. The system suggests learning videos according to the user's knowledge and engagement levels. It employs a sensor-free framework, using the user interaction log for predictions. Depending on both inputs, different categories of videos are considered for the recommendation. For instance, if a user's engagement is low and their knowledge is weak, the likelihood of recommending a video on a different topic is increased.

For an in-depth analysis of affective VRS, we refer to the comprehensive overview in Wang and Zhao (2022). The paper examines and categorizes the state-of-the-art in this field while identifying future research challenges. These challenges encompass the (1) scarcity of realistic high-quality datasets, (2) the integration of existing models with emerging deep-learning techniques, and (3) the adaptation of affective VRS for goals beyond accuracy, such as multi-task recommendations and explainable recommendations.

3.6.2. Context-awareness

Context-aware recommender systems extend traditional recommenders by considering not only items and users but also the specific circumstances of the user when suggesting items (Colombo-Mendoza et al., 2015). These systems can be seen as a type of hybrid recommender, incorporating various factors to generate personalized recommendations. The context in this case refers to a combination of diverse attributes, including *spatial* context (location-related details) and *temporal* context (current time) and their impact on the recommendation process. Context awareness can be introduced to an existing video recommender by filtering or re-ranking its suggestions based on user context (Abbas et al., 2017). By tracking the user's context during video consumption, such as location or time, the system detects different contexts and then removes recommendations that do not align with the user's current context.

Addressing the challenge of identifying suitable contexts for videos watched by diverse users, the usage of *Soft-Rough sets* was proposed in Abbas and Amjad Alam (2019). While traditional *rough sets* handle incomplete or uncertain data by extracting patterns, they struggled to establish decision rules for video-context detection. Soft-rough sets, however, expand on rough sets by incorporating similarity degrees, enabling more flexible data classification and analysis. This extension helps in identifying the most fitting video context. In Abbas et al. (2019), a solution is introduced to address the problem of contextual sparsity in video recommendations, where relevant contexts are scarce due to insufficient data. Existing methods with uniform context weights often conflicted when choosing appropriate contexts for videos. To address this, a soft-rough set-based attribute reduction technique was employed. This technique identifies a minimal influential set of contextual factors that meet users' requirements within the VRS. Recommendations are drawn directly from computed soft sets of videos and contexts, with conflict-free recommendations being straightforward. In cases of conflict, attribute weights are determined by assessing the interdependency of contexts. Attributes that better differentiate contexts receive higher weights, aiding in selecting pertinent contexts for a given video set.

3.6.3. Segments of interest

Segments of Interest (SOI) are video parts that users highlight while watching because they are interesting to them. The intention is that users like specific parts of videos more than others. In Dias et al. (2013), users with overlapping SOIs in different videos are assumed to have similar tastes and are selected as nearest neighbors for video recommendations. The SOI similarity is used to increase the similarity between users with overlaps proportionally, impacting the nearest neighbor computation while avoiding issues when no segments are highlighted yet.

An alternative approach to highlight SOI is introduced in Ferracani et al. (2015). Users annotate outstanding frames with comments and add semantic references to WIKIPEDIA.[12] These annotations are used to cluster the video into a hierarchically structured taxonomy using the fuzzy k-Means algorithm. Videos

12 https://www.wikipedia.org

are represented as vectors of weighted categories, used to determine video similarity. Relevance to users is assessed by merging implicit and explicit ratings.

3.6.4. Datasets

Publicly available datasets are valuable resources for researchers to compare the results of offline experiments and enable reproducibility. This way benchmarks and leaderboards can be created, providing an overview of the state-of-the-art performance in specific domains. In the field of RS, platforms like *Papers With Code*[13] offer benchmarks for various datasets, including those relevant to VRS, fostering accessibility to datasets with diverse characteristics.

One of the most used datasets for RS and especially VRS are the *MovieLens* datasets (Harper and Konstan, 2015). Launched by researchers at the University of Minnesota in 1997, *MovieLens* is a movie recommendation system that allows users to rate movies and receive personalized recommendations based on their ratings. Based on the collected data of this service, multiple versions of the dataset with different sizes have been released over the years, making it a standard benchmark for recommender algorithms in research and education.

The NETFLIX dataset (Bennett and Lanning, 2007), released in 2006 alongside the *Netflix prize* challenge, contains anonymous movie ratings by users. The challenge aimed to outperform the accuracy of the *Cinematch* baseline by 10%, measured using *Root mean squared error (RMSE)* as metric. The goal was to predict the number of stars a user would rate a movie on a 1 to 5 scale. This competition resulted in significant advancements in RS, with matrix factorization methods becoming key technologies for collaborative filtering, surpassing classical nearest-neighbor techniques. The winning solution is detailed in Koren (2009).

The Supplementary material of this paper offers a range of datasets for assessing and enhancing VRS. These datasets are outlined with a short description. Most datasets are suitable for content-based and collaborative filtering, with fewer incorporating context awareness and affective signals. Entertainment domains, particularly movies, dominate the dataset landscape, with fewer options for domains like e-learning, resulting in fewer research publications in those areas. This scarcity of specialized datasets emphasizes the need for more domain-specific datasets to foster research in various areas.

3.6.5. Evaluation metrics

Evaluation metrics are essential in VRS experiments, offering insights into recommendation quality. Consistent metrics across publications enable system comparison and finding suitable approaches. A wide range of metrics assess various quality aspects, including accuracy, coverage, novelty, and scalability, across different item types, including videos. A comprehensive overview of RS evaluation, including offline and online settings, is available in Gunawardana et al. (2022).

In the context of video recommendations, *unexpectedness* was introduced as a unique concept in RS in Adamopoulos and Tuzhilin (2014). Unlike *novelty*, which suggests unfamiliar items, unexpectedness recommends items that deviate from user expectations but are still perceived as beneficial. *Serendipity* goes further, requiring user appreciation for the recommendation and excluding items that are not novel, while unexpectedness may include surprising but known items. *Diversification* enhances item variety through post-processing by removing or replacing similar items, unlike unexpectedness, which affects recommendation generation. Integrating unexpectedness with accuracy can enhance overall user satisfaction. In addition, the *Bayesian Surprise* measures computational creativity by quantifying surprise as the distance between user expectations, aiding the development of creative and surprising recommendations (Lu et al., 2018).

4. Discussion

In recent years, various approaches have been introduced for recommending videos in different situations. Due to the complexity and diversity of applications, there is no single solution that can be universally applied in all contexts. The choice of the appropriate approach depends on specific objectives. Addressing various challenges requires different mitigation strategies, which will be discussed in the following section, and finally, concluded by highlighting potential areas for future research and addressing unresolved issues.

Content-based video recommendation approaches do not rely on user communities and are applicable to individual users by understanding their interests and the available content. These methods suggest videos with content most similar to the user's preferences (Adomavicius and Tuzhilin, 2005; Jannach et al., 2011b; Nikolakopoulos et al., 2022). However, knowledge about user interests is crucial, which can be acquired explicitly through ratings (Lee and Abu-El-Haija, 2017) or direct preferences (Sanchez et al., 2012; Tavakoli et al., 2020), or implicitly through user-system interactions (Mei et al., 2007, 2011; Liu et al., 2020).

A more advanced method for automatically gathering implicit feedback involves the utilization of affective sensors, which is a popular topic of active research. These sensors have the potential to enhance the interpretation of implicit feedback, leading to improved recommendations (Choi et al., 2016; Kaklauskas et al., 2018; Okubo and Tamura, 2019; Kim et al., 2021). However, their widespread adoption faces uncertainty due to user acceptance and privacy concerns, particularly for more complex devices like EEGs (Bandara et al., 2021). Ensuring responsible usage and compliance with privacy laws, such as GDPR[14], is crucial to building user trust in such technologies.

In general, content-based approaches have some common weaknesses (Adomavicius and Tuzhilin, 2005; Nikolakopoulos et al., 2022): (1) *Limited content analysis* arises from incomplete or insufficient information about items and users, hindering personalized recommendations. (2) *Over-specialization* occurs as these approaches mainly focus on suggesting similar items to those previously liked, potentially missing diverse content relevant to the user. (3) The *cold start problem* describes a ramp-up phase of new users to a system, requiring new users to provide enough ratings

13 https://paperswithcode.com/task/recommendation-systems

14 https://gdpr.eu

for the system to generate useful recommendations, which may take time.

To address challenges like the cold start problem and limited content analysis, automatic extraction of features has proven effective in representing video content for recommendation (Luo et al., 2008; Ramezani and Yaghmaee, 2016; Lee and Abu-El-Haija, 2017; Hazrati and Elahi, 2021; Rimaz et al., 2021). The selection of features impacts recommendation quality, with different multimedia features showing varying effectiveness across video domains. For instance, in domains rich in information density like education or news, textual features appear to provide the most valuable content description (Luo et al., 2008; Chantanurak et al., 2016; Kimoto et al., 2016; Tavakoli et al., 2020). In contrast, in entertainment domains, especially visual features appear to offer a good basis for calculation of recommendations (Deldjoo et al., 2016, 2018b; Lee and Abu-El-Haija, 2017; Elahi et al., 2020, 2021; Yi et al., 2022).

Combining multiple features of different types can improve recommendation quality in some cases (Elahi et al., 2017; Deldjoo et al., 2018a). However, this is not universally valid. For instance, combining stylistic visual features with textual content descriptions in the movie domain may reduce quality due to semantic dissimilarity (Deldjoo et al., 2018b). In some cases, using low-level visual features individually outperforms their combination due to the lack of correlation between aspects (Deldjoo et al., 2016). The quality of recommendations also depends on the aggregation strategies used (Mei et al., 2007, 2011; Chakder et al., 2022; Pingali et al., 2022; Mondal et al., 2023), with different contexts requiring different aggregation approaches for better performance.

In Section 3.2, various algorithms with distinct requirements for optimal performance were identified. Supervised learning techniques excel with good feature descriptors, particularly when leveraging textual features (Sanchez et al., 2012; Tavakoli et al., 2020). They work well even with limited user information, making them valuable for new users (Sanchez et al., 2012). Unsupervised techniques perform effectively with sparse feature descriptions, enabling the retrieval of meaningful topic descriptors (Wu et al., 2008; Lu et al., 2017). For entertainment videos, automatically extracted low-level visual features are well-suited for clustering-based recommendations, outperforming manually added textual features (Deldjoo et al., 2016, 2018b). Clustering also helps maintain performance in large item catalogs, as only the most similar clusters to the user profile need consideration. Self-supervised approaches are suitable for large catalogs, especially when used in conjunction with automatically extracted features. Deep neural networks are often applied for CTR prediction to recommend videos the user is likely to watch next (Covington et al., 2016; Liu et al., 2020). Multi-modal features are effective for video representation, capturing hidden commonalities between items and utilizing comprehensive descriptions for robust recommendations (Chakder et al., 2022; Pingali et al., 2022; Mondal et al., 2023).

With the availability of user ratings, collaborative filtering is a widely used technique for video recommendation, especially in scenarios with many users. Unlike content-based approaches, CF does not require content analysis, as long as explicit or implicit ratings are present (Jannach et al., 2011a; Nikolakopoulos et al., 2022). However, CF systems face two kinds of cold start problems: (1) the *new user problem* requires new users to provide enough ratings, and (2) the *new item problem*, where new items require enough ratings to be recommended. Furthermore, the *sparsity* of ratings challenge those systems, as a sufficient number is crucial for accurate recommendations (Adomavicius and Tuzhilin, 2005).

For collaborative filtering in video recommendation, the kNN method is frequently used. Similar users are identified as neighbors based on their rating patterns, and their ratings are used to predict ratings for the target user (Dias et al., 2013). To handle large user datasets and maintain sufficient performance, clustering is applied to focus on relevant data subsets (Katarya and Verma, 2016; Katarya, 2018). To address the sparsity of ratings, implicit feedback is employed to learn preferences from past user interactions (He et al., 2017; Rybakov et al., 2018). Especially self-supervised approaches have demonstrated effectiveness in handling implicit ratings efficiently.

In general, CF approaches are effective in avoiding overspecialization and enhancing recommendation quality in terms of serendipity, regardless of the specific method used. This was demonstrated with the winning system of the *Netflix prize*, which employed matrix factorization techniques (Koren et al., 2009).

To mitigate cold start situations for new users in CF, sharing user information across multiple platforms or social networks can be effective in providing initial user profiles (Deng et al., 2013; Yan et al., 2019). However, its real-life applicability is limited to cases where one provider offers multiple services and can share data between them, with privacy protection being a critical consideration. Alternatively, using demographic information for initial recommendations to new users can be helpful (Cui et al., 2014; Vizine Pereira and Hruschka, 2015), extending CF to a hybrid approach.

Hybrid video recommenders combine different methods to overcome individual limitations. To address cold start for new users, hybrids merge CF with CBF by enriching user profiles from other sources (Cui et al., 2014; Vizine Pereira and Hruschka, 2015) or augmenting items with content descriptions (Öztürk and Kesim Cicekli, 2011; Wang et al., 2015, 2021; Gao et al., 2017; Mahadevan and Arock, 2017; Wei et al., 2017; Liu et al., 2019b; Kvifte et al., 2021). The latter is particularly helpful in mitigating the sparsity of user ratings. Additionally, Multi-Task Learning can be used to effectively combine multiple objectives within a single VRS (Zhao et al., 2019; Tang et al., 2020).

By adding context information to video recommenders, the challenge of changing user interests based on spatial or temporal context can be addressed. These systems incorporate information about when and where users consume videos, allowing them to provide more relevant and useful recommendations, ultimately enhancing the overall user experience (Abbas and Amjad Alam, 2019; Abbas et al., 2019).

As a summary, we conclude our findings in Table 6 by outlining the advantages and disadvantages of the different approaches for video recommendation. While content-based methods serve as a good standard approach for video recommendation when at least basic feature descriptions exist or can be generated, the incorporation of user ratings enables the utilization of collaborative methods, which frequently enhance the generation of unexpected suggestions. However, these methods require a ramp-up phase to be able to suggest useful videos. A hybrid approach that merges content features with collaborative data presents a good opportunity to alleviate the limitations and leverage the advantages of each approach.

In cases, where the recommendation of videos is directed toward multiple persons instead of individuals, group recommender systems are able to suggest content that satisfies the preferences of multiple users simultaneously. The challenge is to balance diverse user profiles and recommend items in a suitable order (O'Connor et al., 2001; Yu et al., 2006). While group recommendation can be beneficial, it is not widely used for videos compared to individual user-based approaches. However, it offers potential advantages, such as more expressive ratings when different criteria are rated separately, to understand why a user likes the video, and compute recommendations based on those criteria (Felfernig et al., 2018; Masthoff and Delić, 2022). Furthermore, cold start situations can be mitigated by using social filtering to extend user profiles with information from similar users.

4.1. Research issues

Our literature overview on video recommender systems highlights several potential research directions for further exploration in this field. These directions will be elaborated on in the following.

4.1.1. Bias and manipulation

Recent attention has been drawn to bias in video recommendations, particularly in social and political contexts, like elections and the COVID-19 pandemic. Platforms like YouTube are accused to steer users in specific directions or causing filter bubbles, and spreading misinformation. Yet, publications analyzing bias in video recommendations are scarce. One such study (Kirdemir et al., 2021) investigated bias in YouTube's algorithm, finding that a few videos are recommended noticeably more frequently, creating a bias toward popular videos. In Papadamou et al. (2022), the recommendation of pseudoscientific content, e.g., videos promoting conspiracy theories, on YouTube was analyzed to observe the self-reinforcing effect of the view history, showing that countermeasures to fight misinformation are part of the recommendation algorithm.

Besides bias, manipulating recommendations is a significant concern explored across various item domains (Hurley, 2011; Adomavicius et al., 2013), particularly on social media platforms (Lang et al., 2010). The study in Edwards et al. (2022) illustrated a successful attack on a content-based recommender using manipulated videos, where subtle modifications to video visual features affected the model's content interpretation, while it was not recognizable to the human eye.

Based on this initial research, improving the understanding and increasing the awareness of bias in video recommendation can be a promising research area. Furthermore, researching methods for detecting and preventing manipulation also presents a potential for future work.

4.1.2. Few-shot and zero-shot video recommendation

Recently, neural network models capable of *few-shot* and *zero-shot* classification, like, for example, *CLIP* (Radford et al., 2021), gained increasing attention. Those models are able to accurately predict labels with few (few-shot) or none (zero-shot) labeled examples. While these models already have been shown to outperform other approaches in interactive video retrieval (Lokoč et al., 2023), their potential in video recommendation remains largely unexplored. Future research could focus on applications in recommendation systems where historical interaction data is limited or absent, potentially improving cold start scenarios. Additionally, the possibility of developing generalized models capable of accurately recommending videos across diverse domains offers potential for future work.

4.1.3. Live stream recommendation

Incorporating recommenders in live stream scenarios presents a promising field with real-time performance requirements. While real-time feedback analysis via affective sensors has been explored (see Section 3.6.1), limited attention has been given to live content analysis. For instance, in Dai et al. (2014), an approach using OCR and figure recognition on keyframes has been proposed to detect text and suggest related videos during live streams, like showing additional videos of a scoring football player. The key challenge involves rapid feature extraction and computation to understand live stream content for timely recommendations. A potential direction for future research could involve exploring various options for applying recommendations in live stream contexts.

4.1.4. Knowledge-based video recommendation

Knowledge-based recommender systems leverage information about items and users to make reasoned decisions about which items align with user requirements in an interactive manner (Burke, 2000; Felfernig and Burke, 2008). Users specify their preferences, and the system attempts to identify suitable items. If none are found, user requirements might need adjustment (Jannach et al., 2011d). While this approach is well-established in various domains, particularly in cases where items are complex or users have limited knowledge about them, e.g., financial services, it remains underexplored for videos. This scarcity of publications might be related to the perceived high cost of defining recommendation knowledge for large video catalogs. However, in domains like learning videos, knowledge-based systems could be beneficial, allowing users to express their knowledge and refine their requirements iteratively, as outlined in Lubos et al. (2022). Users with general learning goals can outline their existing knowledge as requirements, allowing iterative refinement. Case-based systems (Jannach et al., 2011d), which allow users to refine their requirements iteratively, could guide users to appropriate videos. Initial studies in this area can be valuable

TABLE 6 Advantages and disadvantages of different recommendation approaches in the video domain.

Content-based RS		Collaborative Filtering		Hybrid RS	
Advantages	Disadvantages	Advantages	Disadvantages	Advantages	Disadvantages
No user community required	Modeling of content representation	No need for content representation	Sufficiently large user base required	Mitigate cold start for new users	Increased maintenance cost
High scalability	Learning user preferences	Serendipity	Cold start for new items	Mitigate low number of ratings	Computational complexity
No cold start for new items (extracted content features)	Cold start for new users	No explicit modeling of user preferences	Cold start for new users	Extension of user profiles with other sources	
Niche item recommendation	Overspecialization due to focus on similarity	Offline computation		Consideration of user context	

to assess the applicability of knowledge-based approaches for video recommendations.

4.1.5. Multi-modal content representation

Video items are characterized by multi-modality, incorporating various dimensions that describe their content (see Section 3.2). yielding rich information potential yet posing efficiency challenges in representation. While existing studies (Mei et al., 2007, 2011; Chakder et al., 2022; Pingali et al., 2022; Mondal et al., 2023) address this topic, many questions remain unanswered. Future research can focus on the analysis and development of methods to aggregate multi-modal features, across diverse video domains and applications, to determine effective strategies for specific scenarios. Furthermore, a performance comparison between recommenders using aggregated feature descriptions and systems aggregating the suggestions of multiple systems operating on distinct dimensions could be considered. This could help identify effective strategies for content representation and recommendation.

4.1.6. Non-entertainment datasets

Most video recommendation datasets concentrate on the entertainment domain, particularly movies (see Section 3.6.4). This leaves a gap in publicly available datasets from other domains like e-learning, where the content is substantially different. As a result, evaluation outcomes derived from entertainment datasets might not accurately reflect system performance in other scenarios. Given the increasing significance of videos across diverse domains, particularly in knowledge transfer, there is a need for advancing research and introducing new datasets to aid the development of specialized systems.

4.1.7. Scalability

As the demand for personalized video recommendations grows, video streaming companies face challenges related to hardware and network traffic. To ensure a stable service, cloud servers are distributed. However, this can lead to localized biases in recommendations based on user preferences in that area (Duan et al., 2020). For instance, if a local server serves mainly young users who prefer educational content, older users with different interests might receive inappropriate suggestions. Therefore, one potential for further research can be identified in the distribution of RS on cloud and edge infrastructures, facing the challenges of network load and performance to provide good results in general. The *JointRec* framework, presented in Duan et al. (2020), proposes the *JointCloud* architecture in mobile IoT, using distributed training across servers to mitigate biases and provide competitive results. Further research might explore the potential of distributed VRS in cloud and edge infrastructures.

4.1.8. Segment recommendation

Current video recommender systems primarily focus on suggesting complete videos, which is well-suited for entertainment content. However, in domains like news or education, recommending specific video segments can be more advantageous, as users may only be interested in specific parts of the whole video (see Section 3.6.3). For instance, in knowledge transfer, suggesting relevant segments based on a user's existing knowledge can enhance efficiency by avoiding the repetition of known topics. Future research could explore methods to recognize feedback on specific video parts and interpret this feedback to identify segment borders. Additionally, incorporating user knowledge into their profile preferences is crucial for providing valuable recommendations in such scenarios.

5. Conclusion

This article offers a comprehensive overview of recommendation approaches in the video domain. The methodology used in this study analyzed recent publications, categorizing them based on their underlying recommendation approaches. By examining the various systems, we highlighted their respective strengths and weaknesses, providing valuable insights for selecting the most suitable approach for specific application contexts. In this overview, we identified the challenges and opportunities faced by video recommender systems. By improving the understanding of limitations and potential areas of improvement, we aim to inspire further research and development in the field.

Author contributions

SL: Writing—original draft, Writing—review & editing. AF: Writing—review & editing. MT: Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The presented work has been developed within the research project STREAMDIVER which was funded by the Austrian Research Promotion Agency (FFG) under the project number 886205. Supported by TU Graz Open Access Publishing Fund.

Conflict of interest

MT is employed by Streamdiver GmbH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdata.2023.1281614/full#supplementary-material

References

Abbas, M., Riaz, M. U., Rauf, A., Khan, M. T., and Khalid, S. (2017). "Context-aware Youtube recommender system," in *2017 International Conference on Information and Communication Technologies (ICICT)* (Karachi: IEEE), 161–164.

Abbas, S. M., Alam, K. A., and Shamshirband, S. (2019). A soft-rough set based approach for handling contextual sparsity in context-aware video recommender systems. *Mathematics* 7, 740. doi: 10.3390/math7080740

Abbas, S. M., and Amjad Alam, K. (2019). "Exploiting relevant context with soft-rough sets in context-aware video recommender systems," in *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (New Orleans, LA), 1–6.

Adamopoulos, P., and Tuzhilin, A. (2014). On unexpectedness in recommender systems: or how to better expect the unexpected. *ACM Trans. Intell. Syst. Technol.* 5, 1–32. doi: 10.1145/2559952

Adomavicius, G., Bockstedt, J. C., Curley, S. P., and Zhang, J. (2013). Do recommender systems manipulate consumer preferences? A study of anchoring effects. *Inform. Syst. Res.* 24, 956–975. doi: 10.1287/isre.2013.0497

Adomavicius, G., and Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowledge Data Eng.* 17, 734–749. doi: 10.1109/TKDE.2005.99

Arapakis, I., Moshfeghi, Y., Joho, H., Ren, R., Hannah, D., and Jose, J. M. (2009). "Integrating facial expressions into user profiling for the improvement of a multimodal recommender system," in *2009 IEEE International Conference on Multimedia and Expo* (New York, NY), 1440–1443.

Baluja, S., Seth, R., Sivakumar, D., Jing, Y., Yagnik, J., Kumar, S., et al. (2008). "Video suggestion and discovery for youtube: taking random walks through the view graph," in *Proceedings of the 17th International Conference on World Wide Web, WWW '08* (New York, NY: Association for Computing Machinery), 895–904.

Bandara, S. K., Wijesinghe, U. C., Jayalath, B. P., Bandara, S. K., Haddela, P. S., and Wickramasinghe, L. M. (2021). "EEG based neuromarketing recommender system for video commercials," in *2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS)*, 11–16.

Baum, L. E., and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *Ann. Math. Stat.* 37, 1554–1563.

Bennett, J., and Lanning, S. (2007). "The Netflix prize," in *Proceedings of KDD up and Workshop* (New York, NY).

Burke, R. (2000). Knowledge-based recommender systems. *Encyclopedia Lib. Inform. Syst.* 69(Suppl. 32), 175–186.

Chakder, D., Mondal, P., Raj, S., Saha, S., Ghosh, A., and Onoe, N. (2022). "Graph network based approaches for multi-modal movie recommendation system," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 409–414.

Chantanurak, N., Punyabukkana, P., and Suchato, A. (2016). "Video recommender system using textual data: its application on lms and serendipity evaluation," in *2016 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 289–295.

Chen, J., Gong, Z., Li, Y., Zhang, H., Yu, H., Zhu, J., et al. (2022a). Meta-path based neighbors for behavioral target generalization in sequential recommendation. *IEEE Trans. Netw. Sci. Eng.* 9, 1658–1667. doi: 10.1109/TNSE.2022.3149328

Chen, J., Gong, Z., Wang, W., Wang, C., Xu, Z., Lv, J., et al. (2022b). Adversarial caching training: unsupervised inductive network representation learning on large-scale graphs. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 7079–7090. doi: 10.1109/TNNLS.2021.3084195

Chen, J., Li, X., Li, Y., Li, P., Wang, M., Zhang, X., et al. (2022c). A simple yet effective layered loss for pre-training of network embedding. *IEEE Trans. Netw. Sci. Eng.* 9, 1827–1837. doi: 10.1109/TNSE.2022.3153643

Chen, J., Peng, J., Qi, L., Chen, G., and Zhang, W. (2019). "Implicit rating methods based on interest preferences of categories for micro-video recommendation," in *Knowledge Science, Engineering and Management*, eds C. Douligeris, D. Karagiannis, and D. Apostolou (Cham: Springer International Publishing), 371–381.

Chen, M.-H., Teng, C.-H., and Chang, P.-C. (2015). Applying artificial immune systems to collaborative filtering for movie recommendation. *Adv. Eng. Inform.* 29, 830–839. doi: 10.1016/j.aei.2015.04.005

Chen, X., Liu, D., Xiong, Z., and Zha, Z.-J. (2021). Learning and fusing multiple user interest representations for micro-video and movie recommendations. *IEEE Trans. Multimedia* 23, 484–496. doi: 10.1109/TMM.2020.2978618

Chen, X., Liu, D., Zha, Z.-J., Zhou, W., Xiong, Z., and Li, Y. (2018). "Temporal hierarchical attention at category- and item-level for micro-video click-through prediction," in *Proceedings of the 26th ACM International Conference on Multimedia, MM '18* (New York, NY: Association for Computing Machinery), 1146–1153.

Choi, I. Y., Oh, M. G., Kim, J. K., and Ryu, Y. U. (2016). Collaborative filtering with facial expressions for online video recommendation. *Int. J. Inform. Manage.* 36, 397–402. doi: 10.1016/j.ijinfomgt.2016.01.005

Colombo-Mendoza, L. O., Valencia-Garcia, R., Rodriguez-Gonzalez, A., Alor-Hernandez, G., and Samper-Zapater, J. J. (2015). Recommetz: a context-aware knowledge-based mobile recommender system for movie showtimes. *Expert Syst. Appl.* 42, 1202–1222. doi: 10.1016/j.eswa.2014.09.016

Covington, P., Adams, J., and Sargin, E. (2016). "Deep neural networks for youtube recommendations," in *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16* (New York, NY: Association for Computing Machinery), 191–198.

Cui, P., Wang, Z., and Su, Z. (2014). "What videos are similar with you? Learning a common attributed representation for video recommendation," in *Proceedings of the 22nd ACM International Conference on Multimedia, MM '14* (New York, NY: Association for Computing Machinery), 597–606.

Dai, Z., Sheng, G., Honggang, Z., Guang, C., Yongsheng, Z., Jifeng, T., et al. (2014). "A real-time video recommendation system for live programs," in *2014 4th IEEE International Conference on Network Infrastructure and Digital Content*, 498–502.

Davidson, J., Livingston, B., Sampath, D., Liebald, B., Liu, J., Nandy, P., et al. (2010). "The YouTube video recommendation system," in *Proceedings of the Fourth ACM Conference on Recommender Systems - RecSys '10* (Barcelona: ACM Press), 293.

De Pessemier, T., Verlee, D., and Martens, L. (2016). "Enhancing recommender systems for tv by face recognition," in *12th International Conference on Web Information Systems and Technologies (WEBIST 2016)*, 243–250.

Deldjoo, Y. (2020). *Enhancing Video Recommendation Using Multimedia Content*. Cham: Springer International Publishing.

Deldjoo, Y., Constantin, M. G., Eghbal-Zadeh, H., Ionescu, B., Schedl, M., and Cremonesi, P. (2018a). "Audio-visual encoding of multimedia content for enhancing movie recommendations," in *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys '18* (New York, NY: Association for Computing Machinery), 455–459.

Deldjoo, Y., Elahi, M., Cremonesi, P., Garzotto, F., Piazzolla, P., and Quadrana, M. (2016). Content-based video recommendation system based on stylistic visual features. *J. Data Seman.* 5, 99–113. doi: 10.1007/s13740-016-0060-9

Deldjoo, Y., Elahi, M., Quadrana, M., and Cremonesi, P. (2018b). Using visual features based on MPEG-7 and deep learning for movie recommendation. *Int. J. Multimedia Inform. Retriev.* 7, 207–219. doi: 10.1007/s13735-018-0155-1

Deldjoo, Y., Schedl, M., Hidasi, B., Wei, Y., and He, X. (2022). *Multimedia Recommender Systems: Algorithms and Challenges*. New York, NY: Springer US.

Deng, Z., Sang, J., and Xu, C. (2013). "Personalized video recommendation based on cross-platform user modeling," in *2013 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6.

Dias, A. S., Wives, L. K., and Roesler, V. (2013). "Enhancing the accuracy of ratings predictions of video recommender system by segments of interest," in *Proceedings of the 19th Brazilian Symposium on Multimedia and the Web, WebMedia '13* (New York, NY: Association for Computing Machinery), 241–248.

Diaz, Y., Alm, C. O., Nwogu, I., and Bailey, R. (2018). "Towards an affective video recommendation system," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)* (Athens: IEEE), 137–142.

Du, Q., Yu, L., Li, H., Ou, N., Gong, X., and Xiang, J. (2022). "M3REC: cross-modal context enhanced micro-video recommendation with mutual information maximization," in *2022 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6.

Duan, S., Zhang, D., Wang, Y., Li, L., and Zhang, Y. (2020). JointREC: a deep-learning-based joint cloud video recommendation framework for mobile IoT. *IEEE Internet Things J.* 7, 1655–1666. doi: 10.1109/JIOT.2019.2944889

Edwards, D., Rawat, D. B., and Sadler, B. M. (2022). "Adversarial promotion for video based recommender systems," in *2022 IEEE 4th International Conference on Cognitive Machine Intelligence (CogMI)*, 134–138.

Elahi, M., Bakhshandegan Moghaddam, F., Hosseini, R., Rimaz, M. H., El Ioini, N., Tkalcic, M., et al. (2021). *Recommending Videos in Cold Start With Automatic Visual Tags*. New York, NY: Association for Computing Machinery.

Elahi, M., Deldjoo, Y., Bakhshandegan Moghaddam, F., Cella, L., Cereda, S., and Cremonesi, P. (2017). "Exploring the semantic gap for movie recommendations," in *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys '17* (New York, NY: Association for Computing Machinery), 326–330.

Elahi, M., Hosseini, R., Rimaz, M. H., Moghaddam, F. B., and Trattner, C. (2020). "Visually-aware video recommendation in the cold start," in *Proceedings of the 31st ACM Conference on Hypertext and Social Media, HT '20* (New York, NY: Association for Computing Machinery), 225–229.

Felfernig, A., Boratto, L., Stettinger, M., Tkalčič, M., et al. (2018). *Group Recommender Systems: An Introduction*. Springer.

Felfernig, A., and Burke, R. (2008). "Constraint-based recommender systems: technologies and research issues," in *Proceedings of the 10th International Conference on Electronic Commerce, ICEC '08* (New York, NY: Association for Computing Machinery).

Ferracani, A., Pezzatini, D., Bertini, M., Meucci, S., and Del Bimbo, A. (2015). "A system for video recommendation using visual saliency, crowdsourced and automatic annotations," in *Proceedings of the 23rd ACM International Conference on Multimedia, MM '15* (New York, NY: Association for Computing Machinery), 757–758.

Gao, J., Zhang, T., and Xu, C. (2017). "A unified personalized video recommendation via dynamic recurrent neural networks," in *Proceedings of the 25th ACM International Conference on Multimedia, MM '17* (New York, NY: Association for Computing Machinery), 127–135.

Gomez-Uribe, C. A., and Hunt, N. (2016). The netflix recommender system: algorithms, business value, and innovation. *ACM Trans. Manage. Inf. Syst.* 6, 1–19. doi: 10.1145/2843948

Gong, X., Feng, Q., Zhang, Y., Qin, J., Ding, W., Li, B., et al. (2022). "Real-time short video recommendation on mobile devices," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, CIKM '22* (New York, NY: Association for Computing Machinery), 3103–3112.

Gunawardana, A., Shani, G., and Yogev, S. (2022). *Evaluating Recommender Systems*. New York, NY: Springer US.

Harper, F. M., and Konstan, J. A. (2015). The movielens datasets: history and context. *ACM Trans. Interact. Intell. Syst.* 5, 1–19. doi: 10.1145/2827872

Hazrati, N., and Elahi, M. (2021). Addressing the new item problem in video recommender systems by incorporation of visual features with restricted Boltzmann machines. *Expert Syst.* 38. doi: 10.1111/exsy.12645

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., and Chua, T.-S. (2017). "Neural collaborative filtering," in *Proceedings of the 26th International Conference on World Wide Web, WWW '17* (Geneva: International World Wide Web Conferences Steering Committee), 173–182.

Ho, T. K. (1995). "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition* (IEEE), 278–282.

Hongliang, C., and Xiaona, Q. (2015). "The video recommendation system based on DBN," in *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing* (Liverpool: IEEE), 1016–1021.

Hopfgartner, F., and Jose, J. M. (2010). Semantic user profiling techniques for personalised multimedia recommendation. *Multimedia Syst.* 16, 255–274. doi: 10.1007/s00530-010-0189-6

Huang, X., Song, Q., Yang, F., and Hu, X. (2019). "Large-scale heterogeneous feature embedding," in *Proceedings of the AAAI Conference on Artificial Intelligence* (Honolulu, HI: AAAI Press).

Hurley, N. J. (2011). "Robustness of recommender systems," in *Proceedings of the Fifth ACM Conference on Recommender Systems* (Chicago, IL), 9–10.

Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2011a). *Collaborative Recommendation* (Cambridge: Cambridge University Press), 13–50.

Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2011b). *Content-Based Recommendation* (Cambridge: Cambridge University Press), 51–80.

Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2011c). *Hybrid Recommendation Approaches* (Cambridge: Cambridge University Press), 124–142.

Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2011d). *Knowledge-Based Recommendation* (Cambridge: Cambridge University Press), 81–123.

Jayalakshmi, S., Ganesh, N., Čep, R., and Senthil Murugan, J. (2022). Movie recommender systems: concepts, methods, challenges, and future directions. *Sensors* 22. doi: 10.3390/s22134904

Jiang, H., Wang, W., Wei, Y., Gao, Z., Wang, Y., and Nie, L. (2020). "What aspect do you like: multi-scale time-aware user interest modeling for micro-video recommendation," in *Proceedings of the 28th ACM International Conference on Multimedia, MM '20* (New York, NY: Association for Computing Machinery), 3487–3495.

Kaklauskas, A., Zavadskas, E., Banaitis, A., Meidute-Kavaliauskiene, I., Liberman, A., Dzitac, S., et al. (2018). A neuro-advertising property video recommendation system. *Technol. Forecast. Soc. Change* 131, 78–93. doi: 10.1016/j.techfore.2017.07.011

Katarya, R. (2018). Movie recommender system with metaheuristic artificial bee. *Neural Comput. Appl.* 30, 1983–1990. doi: 10.1007/s00521-017-3338-4

Katarya, R., and Verma, O. P. (2016). A collaborative recommender system enhanced with particle swarm optimization technique. *Multimedia Tools Appl.* 75, 9225–9239. doi: 10.1007/s11042-016-3481-4

Kim, G., Choi, I., Li, Q., and Kim, J. (2021). A CNN-based advertisement recommendation through real-time user face recognition. *Appl. Sci.* 11. doi: 10.3390/app11209705

Kimoto, M., Nakahata, T., Hirano, T., Nagashio, T., Shiomi, M., Iio, T., et al. (2016). "Video recommendation system that arranges video clips based on pre-defined viewing times," in *Human Interface and the Management of Information: Applications and Services*, ed S. Yamamoto (Cham. Springer International Publishing), 478–486.

Kirdemir, B., Kready, J., Mead, E., Hussain, M. N., Agarwal, N., and Adjeroh, D. (2021). "Assessing bias in youtube's video recommendation algorithm in a cross-lingual and cross-topical context," in *Social, Cultural, and Behavioral Modeling*, eds R. Thomson, M. N. Hussain, C. Dancy, and A. Pyke (Cham: Springer International Publishing), 71–80.

Koren, Y. (2009). The bellkor solution to the Netflix grand prize. *Netflix Prize Document.* 81, 1–10.

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer* 42, 30–37. doi: 10.1109/MC.2009.263

Koren, Y., Rendle, S., and Bell, R. (2022). *Advances in Collaborative Filtering*. New York, NY: Springer US.

Kvifte, T., Elahi, M., and Trattner, C. (2021). "Hybrid recommendation of movies based on deep content features," in *Service-Oriented Computing–ICSOC 2021 Workshops: AIOps, STRAPS, AI-PA and Satellite Events* (Berlin; Heidelberg: Springer-Verlag), 32–45.

Lang, J., Spear, M., and Wu, S. F. (2010). "Social manipulation of online recommender systems," in *Social Informatics: Second International Conference, SocInfo 2010* (Laxenburg: Springer), 125–139.

Lee, J., and Abu-El-Haija, S. (2017). "Large-scale content-only video recommendation," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 987–995.

Leite, L. W., Roy, S., Chakraborty, N., Michailidis, G., Huggins-Manley, A. C., D'Mello, S., et al. (2022). "A novel video recommendation system for algebra: An effectiveness evaluation study," in *LAK22: 12th International Learning Analytics and Knowledge Conference, LAK22* (New York, NY: Association for Computing Machinery), 294–303.

Li, Y., Liu, M., Yin, J., Cui, C., Xu, X.-S., and Nie, L. (2019). "Routing micro-videos via a temporal graph-guided recommendation system," in *Proceedings of the 27th ACM International Conference on Multimedia, MM '19* (New York, NY: Association for Computing Machinery), 1464–1472.

Liu, B., Hsu, W., and Ma, Y. (1998). "Integrating classification and association rule mining," in *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, KDD'98* (AAAI Press), 80–86.

Liu, Y., Lyu, C., Liu, Z., and Tao, D. (2019a). "Building effective short video recommendation," in *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 651–656.

Liu, Y., Yang, T., and Qi, T. (2020). An attention-based user preference matching network for recommender system. *IEEE Access* 8, 41100–41107. doi: 10.1109/ACCESS.2020.2976455

Liu, Y., Zhang, G., Jin, X., and Yuan, H. (2019b). Multi-info fusion based video recommendation system. *J. Phys. Conf. Ser.* 1229, 012010. doi: 10.1088/1742-6596/1229/1/012010

Lokoč, J., Andreadis, S., Bailer, W., Duane, A., Gurrin, C., Ma, Z., et al. (2023). Interactive video retrieval in the age of effective joint embedding deep models: lessons from the 11th VBS. *Multimedia Syst.* 1–24. doi: 10.1007/s00530-023-01143-5

Lu, W., Chung, F.-L., Jiang, W., Ester, M., and Liu, W. (2018). A deep Bayesian tensor-based system for video recommendation. *ACM Trans. Inf. Syst.* 37, 1–22. doi: 10.1145/3233773

Lu, W., Chung, F.-L., Lai, K., and Zhang, L. (2017). Recommender system based on scarce information mining. *Neural Netw.* 93, 256–266. doi: 10.1016/j.neunet.2017.05.001

Lubos, S., Tautschnig, M., Felfernig, A., and Le, V.-M. (2022). "Knowledge-based configuration of videos using feature models," in *Proceedings of the 26th ACM International Systems and Software Product Line Conference - Volume B, SPLC '22* (New York, NY: Association for Computing Machinery), 188–192.

Luo, H., Fan, J., and Keim, D. A. (2008). "Personalized news video recommendation," in *Proceedings of the 16th ACM International Conference on Multimedia, MM '08* (New York, NY: Association for Computing Machinery), 1001–1002.

Ma, S., Zha, Z., and Wu, F. (2019). "Knowing user better: jointly predicting click-through and playtime for micro-video," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*, 472–477.

Mahadevan, A., and Arock, M. (2017). "Credible user-review incorporated collaborative filtering for video recommendation system," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 375–379.

Masthoff, J., and Delić, A. (2022). *Group Recommender Systems: Beyond Preference Aggregation*. New York, NY: Springer US.

Mei, T., Yang, B., Hua, X.-S., and Li, S. (2011). Contextual video recommendation by multimodal relevance and user feedback. *ACM Trans. Inf. Syst.* 29, 1–24. doi: 10.1145/1961209.1961213

Mei, T., Yang, B., Hua, X.-S., Yang, L., Yang, S.-Q., and Li, S. (2007). "VideoReach: an online video recommendation system," in *Proceedings of the 30th annual international ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '07* (Amsterdam: Association for Computing Machinery), 767–768.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. doi: 10.48550/arXiv.1301.3781

Mondal, P., Chakder, D., Raj, S., Saha, S., and Onoe, N. (2023). "Graph convolutional neural network for multimodal movie recommendation," in *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, SAC '23* (New York, NY: Association for Computing Machinery), 1633–1640.

Ng, A. Y., Jordan, M. I., and Weiss, Y. (2001). "On spectral clustering: analysis and an algorithm," in *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, NIPS'01* (Cambridge, MA: MIT Press), 849–856.

Nikolakopoulos, A. N., Ning, X., Desrosiers, C., and Karypis, G. (2022). *Trust Your Neighbors: A Comprehensive Survey of Neighborhood-Based Methods for Recommender Systems*. New York, NY: Springer US.

Niu, J., Zhao, X., Zhu, L., and Li, H. (2013). Affivir: an affect-based internet video recommendation system. *Neurocomputing* 120, 422–433. doi: 10.1016/j.neucom.2012.07.050

O'Connor, M., Cosley, D., Konstan, J. A., and Riedl, J. (2001). "Polylens: a recommender system for groups of user," in *ECSCW*.

Okubo, M., and Tamura, S. (2019). "A proposal of video evaluation method using facial expression for video recommendation system," in *Human Interface and the Management of Information. Information in Intelligent Systems*, eds S. Yamamoto and H. Mori (Cham: Springer International Publishing), 254–268.

Öztürk, G., and Kesim Cicekli, N. (2011). "A hybrid video recommendation system using a graph-based algorithm," in *Modern Approaches in Applied Intelligence*, eds K. G. Mehrotra, C. K. Mohan, J. C. Oh, P. K. Varshney, and M. Ali (Berlin; Heidelberg: Springer), 406–415.

Papadamou, K., Zannettou, S., Blackburn, J., Cristofaro, E. D., Stringhini, G., and Sirivianos, M. (2022). ""It is just a flu": assessing the effect of watch history on youtube's pseudoscientific video recommendations," in *Proceedings of the International AAAI Conference on Web and Social Media*, 723–734.

Pingali, S., Mondal, P., Chakder, D., Saha, S., and Ghosh, A. (2022). "Towards developing a multi-modal video recommendation system," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., et al. (2021). Learning transferable visual models from natural language supervision. International conference on machine learning (PMLR), 8748–8763.

Ramezani, M., and Yaghmaee, F. (2016). A novel video recommendation system based on efficient retrieval of human actions. *Phys. A Stat. Mech. Appl.* 457, 607–623. doi: 10.1016/j.physa.2016.03.101

Ricci, F., Rokach, L., and Shapira, B. (2015). *Recommender Systems: Introduction and Challenges*. Boston, MA: Springer Science; Business Media. 1–36.

Ricci, F., Rokach, L., Shapira, B., and Kantor, P. B., editors (2011). *Recommender Systems Handbook*. New York, NY: Springer.

Rimaz, M. H., Hosseini, R., Elahi, M., and Moghaddam, F. B. (2021). "Audiolens: audio-aware video recommendation for mitigating new item problem," in *Service-Oriented Computing – ICSOC 2020 Workshops*, eds H. Hacid, F. Outay, H. Y. Paik, A. Alloum, M. Petrocchi, M. R. Bouadjenek, A. Beheshti, X. Liu, and A. Maaradji (Cham: Springer International Publishing), 365–378.

Roy, S., and Guntuku, S. C. (2016). "Latent factor representations for cold-start video recommendation," in *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16* (New York, NY: Association for Computing Machinery), 99–106.

Rybakov, O., Mohan, V., Misra, A., Legrand, S., Joseph, R., Chung, K., et al. (2018). "The effectiveness of a two-layer neural network for recommendations," in *ICLR 2018*. Available online at: https://www.amazon.science/publications/the-effectiveness-of-a-two-layer-neural-network-for-recommendations

Sammut, C., and Webb, G. I. (eds.). (2010). *TF–IDF*. Boston, MA: Springer US.

Sanchez, F., Alduan, M., Alvarez, F., Menendez, J. M., and Baez, O. (2012). Recommender system for sport videos based on user audiovisual consumption. *IEEE Trans. Multimedia* 14, 1546–1557. doi: 10.1109/TMM.2012.2217121

Song, J., Jin, B., Yu, Y., Li, B., Dong, X., Zhuo, W., and Zhou, S. (2023). "Mars: a multi-task ranking model for recommending micro-videos," in *Web and Big Data*, eds B. Li, L. Yue, C. Tao, X. Han, D. Calvanese, and T. Amagasa (Cham: Springer Nature), 199–214.

Tang, H., Liu, J., Zhao, M., and Gong, X. (2020). *Progressive Layered Extraction (PLE): A Novel Multi-Task Learning (MTL) Model for Personalized Recommendations*. New York, NY: Association for Computing Machinery.

Tao, J., and Tan, T. (2005). "Affective computing: a review," in *Affective Computing and Intelligent Interaction*, eds J. Tao, T. Tan, and R. W. Picard (Berlin; Heidelberg: Springer), 981–995.

Tavakoli, M., Hakimov, S., Ewerth, R., and Kismihok, G. (2020). "A recommender system for open educational videos based on skill requirements," in *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)* (Tartu: IEEE), 1–5.

Tohidi, N., and Dadkhah, C. (2020). Improving the performance of video collaborative filtering recommender systems using optimization algorithm. *Int. J. Nonlinear Anal. Appl.* 11, 483–495. doi: 10.22075/ijnaa.2020.19127.2058

Véras, D., Prota, T., Bispo, A., Prudencio, R., and Ferraz, C. (2015). A literature review of recommender systems in the television domain. *Expert Syst. Appl.* 42, 9046–9076. doi: 10.1016/j.eswa.2015.06.052

Vizine Pereira, A. L., and Hruschka, E. R. (2015). Simultaneous co-clustering and learning to address the cold start problem in recommender systems. *Knowledge Based Syst.* 82, 11–19. doi: 10.1016/j.knosys.2015.02.016

Wang, D., and Zhao, X. (2022). Affective video recommender systems: a survey. *Front. Neurosci.* 16, 984404. doi: 10.3389/fnins.2022.984404

Wang, H., Wang, N., and Yeung, D.-Y. (2015). "Collaborative deep learning for recommender systems," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15* (New York, NY: Association for Computing Machinery), 1235–1244.

Wang, J., Yessenalina, A., and Roshan-Ghias, A. (2021). "Exploring heterogeneous metadata for video recommendation with two-tower model," in *RecSys 2021 Workshop on Context-Aware Recommender Systems*. Available online at:

https://www.amazon.science/publications/exploring-heterogeneous-metadata-for-video-recommendation-with-two-tower-model

Wang, Z., Yu, X., Feng, N., and Wang, Z. (2014). An improved collaborative movie recommendation system using computational intelligence. *J. Visual Lang. Comput.* 25, 667–675. doi: 10.1016/j.jvlc.2014.09.011

Wei, J., He, J., Chen, K., Zhou, Y., and Tang, Z. (2017). Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Syst. Appl.* 69, 29–39. doi: 10.1016/j.eswa.2016.09.040

Wu, S., Rizoiu, M.-A., and Xie, L. (2019). Estimating attention flow in online video networks. *Proc. ACM Hum. Comput. Interact.* 3, 1–25. doi: 10.1145/3359285

Wu, X., Zhang, Y., Guo, J., and Li, J. (2008). "Web video recommendation and long tail discovering," in *2008 IEEE International Conference on Multimedia and Expo*, 369–372.

Xiao, X., Dai, H., Dong, Q., Niu, S., Liu, Y., and Liu, P. (2023). Social4rec: distilling user preference from social graph for video recommendation in Tencent. *arXiv. [Preprint].arXiv*:2302.09971. Available online at: https://api.semanticscholar.org/CorpusID:257038928

Xue, T., Jin, B., Li, B., Wang, W., Zhang, Q., and Tian, S. (2019). "A spatio-temporal recommender system for on-demand cinemas," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19* (New York, NY: Association for Computing Machinery), 1553–1562.

Yan, H., Chen, X., Gao, C., Li, Y., and Jin, D. (2019). "DeepAPF: deep attentive probabilistic factorization for multi-site video recommendation," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19* (Macao), 1459–1465.

Yi, Z., Wang, X., Ounis, I., and Macdonald, C. (2022). "Multi-modal graph contrastive learning for micro-video recommendation," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22* (New York, NY: Association for Computing Machinery), 1807–1811.

Yu, Z., Zhou, X., Hao, Y., and Gu, J. (2006). Tv program recommendation for multiple viewers based on user profile merging. *User Model. User Adapt. Interact.* 16, 63–82. doi: 10.1007/s11257-006-9005-6

Zhao, Z., Hong, L., Wei, L., Chen, J., Nath, A., Andrews, S., et al. (2019). "Recommending what video to watch next: a multitask ranking system," in *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys '19* (New York, NY: Association for Computing Machinery), 43–51.

Zhou, G., Song, C., Zhu, X., Fan, Y., Zhu, H., Ma, X., et al. (2018). "Deep interest network for click-through rate prediction," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1059–1068.

Zhu, H., Liu, Y., Tian, F., Ni, Y., Wu, K., Chen, Y., and Zheng, Q. (2018). A cross-curriculum video recommendation algorithm based on a video-associated knowledge map. *IEEE Access* 6, 57562–57571. doi: 10.1109/ACCESS.2018.2873106

Zhu, Q., Shyu, M.-L., and Wang, H. (2013). "Videotopic: content-based video recommendation using a topic model," in *2013 IEEE International Symposium on Multimedia*, 219–222.

Zhuo, W., Liu, K., Xue, T., Jin, B., Li, B., Dong, X., et al. (2021). "A behavior-aware graph convolution network model for video recommendation," in *Web and Big Data*, M. Spaniol, Y. Sakurai and J. Chen (Cham: Springer International Publishing), 279–294.

frontiers | Frontiers in Big Data

TYPE Mini Review
PUBLISHED 19 December 2023
DOI 10.3389/fdata.2023.1251072

Check for updates

OPEN ACCESS

EDITED BY
Tyler Derr,
Vanderbilt University, United States

REVIEWED BY
Shaohua Tao,
Xuchang University, China
Sead Delalic,
University of Sarajevo, Bosnia and Herzegovina
Yuying Zhao,
Vanderbilt University, United States

*CORRESPONDENCE
Tomislav Duricic
✉ tduricic@tugraz.at
Dominik Kowald
✉ dkowald@tugraz.at
Elisabeth Lex
✉ elisabeth.lex@tugraz.at

RECEIVED 30 June 2023
ACCEPTED 29 November 2023
PUBLISHED 19 December 2023

CITATION
Duricic T, Kowald D, Lacic E and Lex E (2023) Beyond-accuracy: a review on diversity, serendipity, and fairness in recommender systems based on graph neural networks. *Front. Big Data* 6:1251072.
doi: 10.3389/fdata.2023.1251072

COPYRIGHT
© 2023 Duricic, Kowald, Lacic and Lex. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Beyond-accuracy: a review on diversity, serendipity, and fairness in recommender systems based on graph neural networks

Tomislav Duricic[1,2]*, Dominik Kowald[1,2]*, Emanuel Lacic[3] and Elisabeth Lex[1]*

[1]Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, Austria, [2]Know Center, Graz, Austria, [3]Infobip, Zagreb, Croatia

By providing personalized suggestions to users, recommender systems have become essential to numerous online platforms. Collaborative filtering, particularly graph-based approaches using Graph Neural Networks (GNNs), have demonstrated great results in terms of recommendation accuracy. However, accuracy may not always be the most important criterion for evaluating recommender systems' performance, since beyond-accuracy aspects such as recommendation diversity, serendipity, and fairness can strongly influence user engagement and satisfaction. This review paper focuses on addressing these dimensions in GNN-based recommender systems, going beyond the conventional accuracy-centric perspective. We begin by reviewing recent developments in approaches that improve not only the accuracy-diversity trade-off but also promote serendipity, and fairness in GNN-based recommender systems. We discuss different stages of model development including data preprocessing, graph construction, embedding initialization, propagation layers, embedding fusion, score computation, and training methodologies. Furthermore, we present a look into the practical difficulties encountered in assuring diversity, serendipity, and fairness, while retaining high accuracy. Finally, we discuss potential future research directions for developing more robust GNN-based recommender systems that go beyond the unidimensional perspective of focusing solely on accuracy. This review aims to provide researchers and practitioners with an in-depth understanding of the multifaceted issues that arise when designing GNN-based recommender systems, setting our work apart by offering a comprehensive exploration of beyond-accuracy dimensions.

KEYWORDS
survey, recommender systems, graph neural networks, beyond-accuracy, diversity, serendipity, novelty, fairness

1 Introduction

With their ability to provide personalized suggestions, recommender systems have become an integral part of numerous online platforms by helping users find relevant products and content (Aggarwal et al., 2016). There are various methods employed to implement recommender systems, among which collaborative filtering (CF) has proven to be particularly effective due to its ability to leverage user-item interaction data to generate personalized recommendations (Koren et al., 2021). Recent advances in Graph Neural Networks (GNNs) have also had a significant impact on the field of recommender systems,

10.3389/fdata.2023.1251072

and especially on collaborative filtering. GNN-based CF approaches have demonstrated exceptional results in terms of recommendation accuracy, which has traditionally been the main criterion for evaluating the performance of recommender systems (Pu et al., 2012; He et al., 2020).

However, most studies have focused only on accuracy and have often neglected other equally or sometimes even more important aspects of recommender systems, such as diversity, serendipity, and fairness. The importance of these *beyond-accuracy* dimensions is increasingly being recognized, as studies have shown that these aspects can have a significant impact on user satisfaction (Abdollahpouri et al., 2019). For example, diverse and serendipitous recommendations can prevent the over-specialization of content and enhance user discovery. Novelty, a closely related concept to serendipity, introduces fresh and unexpected options to users, further enriching the discovery process. Fairness, on the other hand, ensures that the system does not discriminate against certain users or item providers, thereby promoting equitable user experiences (Gao et al., 2023).

This review paper further explores these dimensions in the context of GNN-based recommender systems, going beyond the traditional accuracy-centric viewpoint. We discuss recent advances in approaches that not only improve the accuracy-diversity trade-off, but also promote serendipity, novelty and fairness. Furthermore, we highlight the practical issues encountered in assuring these dimensions when constructing GNN-based CF approaches, while preserving high recommendation accuracy. This review is intended to provide researchers and practitioners with a comprehensive understanding of the multifaceted optimization issues that arise when designing GNN-based recommender systems, thereby contributing to the development of more robust and user-centric recommender systems.

2 Background

Graph neural networks (GNNs) have recently emerged as an effective way to learn from graph-structured data by capturing complex patterns and relationships (Hamilton, 2020). Through the propagation and transformation of feature information among interconnected nodes in a graph, GNNs can effectively capture the local and global structure of the given graphs. Consequently, they emerge as an ideal method especially suitable for dealing with tasks involving interconnected, relational data such as social network analysis, molecular chemistry, and recommender systems among others.

In recommender systems, integrating Graph Neural Networks (GNNs) with traditional collaborative filtering techniques has been shown beneficial. Representing users and items as nodes in a graph with interactions acting as edges allows GNNs to provide more accurate personalized recommendations by discovering and utilizing intricate connections that would otherwise remain undetected (Wang X. et al., 2019). In particular, higher-order connectivity together with transitive relationships play an essential role when trying to extract user preferences in certain scenarios.

GNN-based recommender systems represent an evolving field with continuous advancements and innovations. Recent research has focused on multiple aspects of GNNs in recommender systems, ranging from optimizing propagation layers to effectively managing large-scale graphs and integration of auxiliary information (Zhou et al., 2022). Aside from these aspects, an expanding interest lies in exploring beyond-accuracy objectives for recommender systems. Such objectives include diversity, explainability/interpretability, fairness, serendipity/novelty, privacy/security, and robustness which offer a more comprehensive evaluation of the system's performance (Wu S. et al., 2022; Gao et al., 2023). However, our work focuses primarily on three key aspects: diversity, serendipity, and fairness, since these aspects have a significant impact on user satisfaction, while also considering ethical concerns in the field of recommender systems. Ensuring diversity amongst recommendations minimizes over-specialization effects, benefiting users in product/content discovery and exploration (Kunaver and Požrl, 2017). Considering serendipity and novelty also helps to overcome the over-specialization problem by allowing the system to recommend novel and unexpected yet relevant items, thus improving user satisfaction (Kaminskas and Bridge, 2016). The aspect of fairness ensures that the system does not discriminate against certain users or item providers, thereby promoting equitable user experiences (Deldjoo et al., 2023).

Diversity, serendipity, novelty, and fairness in recommender systems are interconnected and often influence each other. For instance, increasing diversity can lead to more serendipitous and novel recommendations, since users are exposed to a wider range of unexpected and less-known items (Kotkov et al., 2020). Some studies occasionally use the terms "diversity" and "novelty" interchangeably, highlighting a common overlap in their conceptual usage (Sun et al., 2020; Dhawan et al., 2022). It's important to note that novelty and serendipity are closer related concepts, as they both compare the recommended items with a user's history, emphasizing the discovery of unexpected content that aligns with personal preferences. Furthermore, focusing on diversity and serendipity can also promote fairness, since it ensures a more equitable distribution of recommendations across items and prevents the system from consistently suggesting only popular items (Mansoury et al., 2020). However, it's important to note that these aspects need to be balanced with the system's accuracy and relevance to maintain user satisfaction. Considering beyond-accuracy dimensions contributes to supporting the development of GNN-based recommender systems that are not only robust and accurate but also user-centric and ethically considerate.

While GNNs have seen rapid advancements, their application in recommender systems has also been the subject of several surveys. Wu S. et al. (2022) and Gao et al. (2023) provide a broad overview of GNN methods in recommender systems, touching upon aspects of diversity and fairness. Dai et al. (2022) delves into fairness in graph neural networks in general, briefly discussing fairness in GNN-based recommender systems. Meanwhile, Fu et al. (2023) explores serendipity in deep learning recommender systems, with limited focus on GNN-based recommenders. Building on these insights, our review distinctively emphasizes the importance of diversity, serendipity, novelty, and fairness in GNN-based recommender systems, offering a deeper dive into these dimensions.

To conduct our review, we searched for literature on Google Scholar using keywords such as "diversity", "serendipity", "novelty", "fairness", "beyond-accuracy", "graph neural networks"

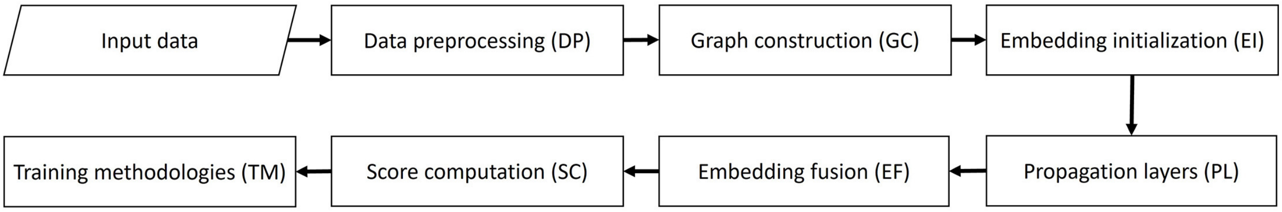

FIGURE 1
The simplified multi-stage process of developing a GNN-based recommender system, each of these stages strongly impacts resulting recommendations and can be considered when designing a model that takes into account beyond-accuracy objectives.

or "recommender system". We manually checked the resulting papers for their relevance and retrieved 20 publications overall from relevant journals and conferences in the field (see Table 1). While re-ranking and post-processing methods are often used when optimizing beyond-accuracy metrics in recommender systems (Gao et al., 2023), this paper specifically concentrates on advancements within GNN-based models, thus leaving these methods outside the discussion. Finally, it is important to highlight that diversity, serendipity, and fairness are extensively researched in recommender systems beyond GNNs. Broader literature across various architectures has provided insights into these challenges and their overarching solutions. While our paper primarily focuses on GNN-based recommender systems, we direct readers to consult these works for a comprehensive perspective (Kaminskas and Bridge, 2016; Castells et al., 2021; Li et al., 2022; Dong et al., 2023; Wang et al., 2023a; Zhao et al., 2023).

3 Model development

The construction of a GNN-based recommender system is a complex, multi-stage process that requires careful planning and execution at each step. These stages include data preprocessing (DP), graph construction (GC), embedding initialization (EI), propagation layers (PL), embedding fusion (EF), score computation (SC), and training methodologies (TM). In this section, we provide an overview of this multi-stage process as it is crucial for understanding the specific stages at which current research has concentrated efforts to address the beyond-accuracy aspects of diversity, serendipity, and fairness in GNN-based recommender systems, as shown in Figure 1.

3.1 Data preprocessing, graph construction, embedding initialization

The initial stage of developing a GNN-based collaborative filtering model is data preprocessing, where user-item interaction data and auxiliary information such as user/item features or social connections are collected and processed (Lacic et al., 2015a; Duricic et al., 2018, 2020; Fan et al., 2019b; Wang H. et al., 2019). Techniques like data imputation ensure that missing data is filled, providing a more complete dataset, while outlier detection helps in maintaining the data's integrity. Feature normalization ensures consistent data scales, enhancing model performance. Addressing the cold-start problem at this stage ensures that new users or items without sufficient interaction history can still receive meaningful recommendations (Lacic et al., 2015b; Liu et al., 2020).

The graph construction stage is crucial, as the graph's structure directly influences the model's efficacy. Choosing the type of graph determines the nature of relationships between nodes. Adjusting edge weights can prioritize certain interactions, while adding virtual nodes/edges can introduce auxiliary information to improve recommendation quality (Kim et al., 2022; Wang et al., 2023b).

In the embedding initialization stage, nodes are assigned low-dimensional vectors or embeddings. The choice of embedding size balances computational efficiency and representation power. Different initialization methods offer trade-offs between convergence speed and stability. Including diverse information in the embeddings can capture richer user-item relationships, enhancing recommendation quality Wang et al. (2021). This initialization can be represented as $H^{(0)} = \left[h^{(0)}_{\text{user}}; h^{(0)}_{\text{item}}\right]$, where $h^{(0)}_{\text{user}}$ and $h^{(0)}_{\text{item}}$ are the initial embeddings of the user and item nodes, respectively.

3.2 Propagation layers, embedding fusion, score computation, training methodologies

Propagation layers in GNNs aggregate and transform features of neighboring nodes to generate node embeddings, represented as $H^{(l+1)} = \sigma\left(D^{-1}AH^{(l)}W^{(l)}\right)$, where $H^{(l)}$ is the matrix of node features at layer l, A is the adjacency matrix, D is the degree matrix, $W^{(l)}$ is the weight matrix at layer l, and σ is the activation function (Hamilton, 2020). There are numerous approaches built on this concept. For instance, He et al. (2020) adopt a simplified approach, emphasizing straightforward neighborhood aggregation to enhance the quality of node embeddings; whereas Fan et al. (2019b) integrate user-item interactions with user-user and item-item relations, capturing complex interactions through a comprehensive graph structure.

Afterward, these embeddings are combined during the embedding fusion stage, forming a latent user-item representation used for score computation by applying a weighted summation, concatenation, or a more complex method of combining user and item embeddings (Wang X. et al., 2019; He et al., 2020).

TABLE 1 This table summarizes key literature on GNN-based recommender systems, emphasizing beyond-accuracy metrics: diversity, serendipity, novelty, and fairness.

Beyond-accuracy goal	References and venue/ journal	Topic/ contribution	Model development stages utilized to tackle metric	Metric
Diversity	Isufi et al. (2021) Information Processing and Management	Neighbor-based mechanism	GC, PL, EF, TM	C, PD
	Ye et al. (2021) ACM RecSys conf.	Dynamic graph construction	EI, GC, TM	PD
	Yang L. et al. (2023) ACM WSDM conf.	Neighbor-based mechanisms	PL, EF, TM	C
	Zuo et al. (2023) MDPI Applied Sciences	Adversarial learning	GC, PL, TM	C, PD
	Ma et al. (2022) IEEE IJCNN conf.	Contrastive learning	EI, GC, PL, TM	C, PD
	Zheng et al. (2021) ACM Web Conf.	Neighbor-based mechanism, Adversarial learning	PL, TM	C, E, GC
	Xie et al. (2021) IEEE Trans. on Big Data	Heterogeneous GNNs	GC, PL, SC, TM	C, LTR, NOV
Serendipity/Novelty	Dhawan et al. (2022) Electronic Commerce Research and Applications	General GNN architecture enhancements	-	SRDP, NOV
	Sun et al. (2020) ACM SIGKDD conf.	General GNN architecture enhancements	GC, PL, EF, SC, TM	SRDP, NOV
	Zhao et al. (2022) ACM SIGIR conf.	Normalization techniques	PL	NOV
	Boo et al. (2023) ACM IUI conf.	Neighbor-based mechanisms	EI, EF, SC, TM	SRDP
Fairness	Xu et al. (2023) Information Sciences	Contrastive learning	GC, TM	ARP
	Li et al. (2019) ACM CIKM conf.	Multimodal feature learning	GC, PL, EF	LTR
	Liu et al. (2022a) Applied Soft Computing	Self-training mechanisms	PL, TM	GF
	Kim et al. (2022) ACM CIKM conf.	Neighbor-based mechanisms	PL, SC, TM	LTR
	Yang Y. et al. (2023) ACM Web Conf.	Contrastive learning	GC, PL, EF, TM	LTR
	Wu K. et al. (2022) ACM ASONAM conf.	Neighbor-based mechanisms	GC, PL, EF, TM	GF
	Gupta et al. (2019) ACM CIKM conf.	Long-tail recommendations	PL, SC, TM	ARP
	Liu and Zheng (2020) ACM RecSys conf.	Long-tail recommendations	DP, EF, SC, TM	C, LTR
	Liu et al. (2022b) Neural Computing and Applications	Neighbor-based mechanisms, Adversarial learning	GC, PL, TM	GF

Each entry specifies the paper's publication venue/journal, a broad strategy categorization, and the model development stages the method utilizes or adapts to enhance the respective metric, including data preprocessing (DP), graph construction (GC), embedding initialization (EI), propagation layers (PL), embedding fusion (EF), score computation (SC), and training methodologies (TM). Additionally, the table highlights which concrete metrics were assessed: Coverage (C), Gini Coefficient (GC), Entropy (E), Pairwise dissimilarity (PD) for Diversity; Serendipity (SRDP); Novelty (NOV); Average Recommendation Popularity (ARP), Group Fairness (GF), and Long Tail Recommendation (LTR) for Fairness.

The score computation stage involves a scoring function to output a score for each user-item pair based on the fused embeddings. The scoring function can be as simple as a dot product between user and item embeddings, or it can be a more complex function that takes into account additional factors (Wang X. et al., 2019; He et al., 2020).

Finally, in the training methodologies stage, a suitable loss function is selected, and an optimization algorithm, typically a variant of stochastic gradient descent, is used to update model parameters (Rendle et al., 2012; Fan et al., 2019a).

Understanding the unique strengths of each stage outlined in this section is essential, and a comparative evaluation can guide the selection of the most suitable approach for specific collaborative filtering scenarios, such as addressing the challenges associated with beyond-accuracy metrics. In Table 1, we provide a comprehensive overview of existing literature, aiding readers in navigating the diverse methodologies and findings discussed throughout this review.

4 Diversity in GNN-based recommender systems

4.1 Definition and importance of diversity

Diversity in recommender systems is a measure of the dissimilarity among the set of items recommended to a user. It prevents over-specialization and enhances user discovery, exposing users to a broader range of items and potentially increasing satisfaction and engagement with the system (Kunaver and Požrl, 2017; Duricic et al., 2021). Diversity can be intra-list, referring to variety within a single recommendation list, or inter-list, concerning variety across different users' lists (Kaminskas and Bridge, 2016). When items are categorized, diversity also entails ensuring a balanced representation of different categories in the recommendations.

Common metrics for measuring diversity include Item Coverage, calculated as the ratio of unique items recommended to the total items in the catalog. The Gini Coefficient reflects recommendation inequality and is given by:

$$\text{Gini Coefficient} = 1 - \sum_{i=1}^{n} P_i^2 \tag{1}$$

where P_i is the proportion of recommendations for item i. Entropy measures unpredictability or randomness in recommendations and is computed as:

$$\text{Entropy} = -\sum_{i=1}^{n} P_i \log P_i \tag{2}$$

with P_i as the probability of item i being recommended (Zheng et al., 2021). Another important metric, Pairwise Dissimilarity, quantifies the average dissimilarity between all pairs of items in a recommendation list (Chen et al., 2018). It is calculated using the formula:

$$\text{Pairwise Dissimilarity} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d(i,j) \tag{3}$$

where N is the number of items in the recommendation list, and $d(i,j)$ represents the measure of dissimilarity between item i and item j.

4.2 Review of recent developments in improving accuracy-diversity trade-off

Several approaches have emerged recently to tackle recommendation diversity using graph neural networks (GNNs). These methods can be broadly categorized based on the specific mechanisms or strategies they employ:

- **Neighbor-based mechanisms[1]:** An approach introduced by Isufi et al. (2021) combines nearest neighbors (NN) and furthest neighbors (FN) with a joint convolutional framework. The *DGRec* method diversifies embedding generation through submodular neighbor selection, layer attention, and loss reweighting (Yang L. et al., 2023). Additionally, *DGCN* model leverages graph convolutional networks for capturing collaborative effects in the user-item bipartite graph, ensuring diverse recommendations through rebalanced neighbor discovery (Zheng et al., 2021).
- **Dynamic graph construction[2]:** *DDGraph* approach involves dynamically constructing a user-item graph to capture both user-item interactions and non-interactions, and then applying a novel candidate item selection operator to choose items from different sub-regions based on distance metrics (Ye et al., 2021).
- **Adversarial learning[3]:** To improve the accuracy-diversity trade-off in tag-aware systems, the *DTGCF* model utilizes personalized category-boosted negative sampling, adversarial learning for category-free embeddings, and specialized regularization techniques (Zuo et al., 2023). Furthermore, the above-mentioned *DGCN* model also employs adversarial learning to make item representations more category-independent.
- **Contrastive learning[4]:** The Contrastive Co-training (*CCT*) method by Ma et al. (2022) employs an iterative pipeline that augments recommendation and contrastive graph views with pseudo edges, leveraging diversified contrastive learning to address popularity and category biases in recommendations.
- **Heterogeneous graph neural networks[5]:** The *GraphDR* approach by Xie et al. (2021) utilizes a heterogeneous graph

1 Neighbor-based mechanisms aggregate and propagate information from neighboring nodes (users or items) to enhance the representation of a target node, capturing intricate relational patterns for improved recommendations (Wu S. et al., 2022).

2 Dynamic graph construction involves continuously updating and evolving the graph structure to incorporate new interactions and/or entities (Skarding et al., 2021).

3 Adversarial examples in recommender systems, as a form of data augmentation, bolster data diversity for improved generalization, counteract inherent biases, and ensure fair node representation in GNNs for fairer recommendations (Deldjoo et al., 2021).

4 Contrastive learning pushes similar item or user embeddings closer and dissimilar ones apart to enhance recommendation quality (Liu et al., 2021).

5 Heterogeneous graph neural networks process diverse types of nodes and edges, capturing complex relationships using a heterogeneous graph as input (Wu S. et al., 2022).

neural network, capturing diverse interactions and prioritizing diversity in the matching module.

Each of these methods offers a unique approach to the accuracy-diversity challenge. While all aim to improve the trade-off, their strategies vary, highlighting the multifaceted nature of the challenge at hand.

5 Serendipity in GNN-based recommender systems

5.1 Definition and importance of serendipity and novelty

Serendipity and novelty are key aspects of recommender systems, essential for enhancing user discovery and engagement. These concepts are closely related and often evaluated together, as they complement each other by simultaneously assessing the unexpectedness and unfamiliarity of recommendations (Sun et al., 2020; Dhawan et al., 2022). Serendipity, indicating the unexpected nature of recommendations, encourages users to explore beyond their usual preferences and stimulates curiosity. The Serendipity Score, is a commonly used metric to assess this quality (Silveira et al., 2019):

$$\text{Serendipity} = \frac{1}{|U|} \sum_{u \in U} \left(\frac{1}{|I_k(u)|} \sum_{i \in I_k(u)} \max(P_i(u) - P_i(U), 0) \cdot rel_i(u) \right) \tag{4}$$

where $|U|$ denotes the cardinality of the user set, $I_k(u)$ the set of top k recommendations for user u, and $rel_i(u)$ the relevance of item i to user u. The difference $P_i(u) - P_i(U)$ captures the preference deviation of user u for item i from the mean user preference.

Conversely, novelty is concerned with how the recommended items are new or unfamiliar to a user, as quantified by the Novelty Score (Zhou et al., 2010):

$$\text{Novelty} = \frac{1}{|U|} \sum_{u \in U} \left(\sum_{i \in I_u(k)} \frac{-\log_2 D(i)}{|I_u(k)|} \right) \tag{5}$$

Here, $D(i)$ signifies the popularity of item i, inversely related to novelty. This measure ensures that recommendations are not only serendipitous but also novel, thus preventing recommendation over-specialization, enhancing user exploration and engagement (Kaminskas and Bridge, 2016).

5.2 Review of recent developments in promoting serendipity and novelty

Recent advancements in GNN-based recommender systems have shown promising results in promoting serendipity and novelty, although notably fewer efforts have been directed toward balancing the accuracy-serendipity and accuracy-novelty trade-offs in comparison to the accuracy-diversity trade-off. In our exploration, we identified several studies addressing these efforts and have categorized them based on the primary theme of their contribution:

- **Neighbor-based mechanisms:** Approach proposed by Boo et al. (2023) enhances session-based recommendations by incorporating serendipitous session embeddings, leveraging session data and user preferences to amplify global embedding effects, enabling users to control explore-exploit tradeoffs.
- **Normalization techniques[6]:** Zhao et al. (2022) proposed *r-AdjNorm*, a simple and effective GNN improvement that can improve the accuracy-novelty trade-off by controlling the normalization strength in the neighborhood aggregation process.
- **General GNN architecture enhancements[7]:** Similarly to the popular *LightGCN* approach by He et al. (2020), the *ImprovedGCN* model by Dhawan et al. (2022) adapts and simplifies the graph convolution process in GCNs for item recommendation, inadvertently boosting serendipity. On the other hand, the *BGCF* framework by Sun et al. (2020), designed for diverse and accurate recommendations, also boosts serendipity and novelty through its joint training approach. These GNN-based models, while focusing on accuracy, inadvertently elevate recommendation serendipity and/or novelty.

These studies collectively demonstrate the potential of GNNs in enhancing the serendipity and novelty of recommender systems, while also highlighting the need for further research to address existing challenges.

6 Fairness in GNN-based recommender systems

6.1 Definition and importance of fairness

Fairness in recommender systems ensures no bias toward certain users or items. It can be divided into user fairness, which avoids algorithmic bias among users or demographics, and item fairness, which ensures equal exposure for items, countering popularity bias (Leonhardt et al., 2018; Kowald et al., 2020; Lex et al., 2020; Abdollahpouri et al., 2021; Lacic et al., 2022). Fairness helps to mitigate bias, supports diversity, and boosts user satisfaction. In GNN-based systems, which can amplify bias, fairness is crucial for balanced recommendations and optimal performance (Ekstrand et al., 2018; Chizari et al., 2022; Chen et al., 2023; Gao et al., 2023).

Key metrics for evaluating fairness include Average Recommendation Popularity (ARP) and Group Fairness (GF) (Yin

6 Normalization techniques in GNN-based recommender systems stabilize and scale node features or edge weights, ensuring consistent and improved model convergence and recommendation quality (Gupta et al., 2019).

7 We refer to general GNN architecture enhancements in recommender systems as the advancements in architectures, aggregators, or training procedures that better capture graph structures for improved recommendation accuracy.

et al., 2012; Fu et al., 2020). ARP, as defined below, assesses the tendency toward recommending popular items:

$$\mathrm{ARP} = \frac{1}{|U|}\sum_{u\in U}\frac{1}{|I_u|}\sum_{i\in I_u} D(i)$$

where $D(i)$ is the popularity of item i, typically defined by the number of interactions or ratings it has received across the user base. On the other hand, GF measures the fairness of recommendations across different user groups:

$$\mathrm{GF} = \left|\frac{1}{|S_0|}\sum_{u\in S_0}\mathcal{T}(Q_u) - \frac{1}{|S_1|}\sum_{u\in S_1}\mathcal{T}(Q_u)\right|$$

Here, S_0 and S_1 represent different user groups, Q_u denotes the list of items recommended to user u, and $\mathcal{T}(Q_u)$ is a metric that scores the quality of recommendations for user u. Lower GF values signify a fairer distribution of recommendations between the groups.

Beyond these metrics, focusing on the assessment of long-tail item recommendations also plays a role in ensuring that the system's suggestions are not limited to well-known or popular items, thus fostering a more inclusive recommendation environment.

6.2 Review of recent developments in promoting fairness

In the evolving landscape of GNN-based recommender systems, the pursuit of user and item fairness has become a prominent topic. Recent advancements can be broadly categorized based on the thematic emphasis of their contributions:

- **Neighbor-based mechanisms:** The *Navip* method debiases the neighbor aggregation process in GNNs using "neighbor aggregation via inverse propensity", focusing on user fairness (Kim et al., 2022). Additionally, the *UGRec* framework by Liu et al. (2022b) employs an information aggregation component and a multihop mechanism to aggregate information from users' higher-order neighbors, ensuring user fairness by considering male and female discrimination. The *SKIPHOP* approach focuses on user fairness by introducing an approach that captures both direct user-item interactions and latent knowledge graph interests, capturing both first-order and second-order proximity. Using fairness for regularization, it ensures balanced recommendations for users with similar profiles (Wu K. et al., 2022).
- **Multimodal feature learning[8]:** The method proposed by Li et al. (2019) fuses hashtag embeddings with multi-modal features, considering interactions among users, micro-videos, and hashtags.
- **Adversarial learning:** The *UGRec* model additionally incorporates adversarial learning to eliminate gender-specific features while preserving common features.
- **Contrastive learning:** The *DCRec* model by Yang Y. et al. (2023) leverages debiased contrastive learning to counteract popularity bias and addressing the challenge of disentangling user conformity from genuine interest, focusing on user fairness. The *TAGCL* framework also capitalizes on the contrastive learning paradigm, ensuring item fairness by reducing biases in social tagging systems (Xu et al., 2023).
- **Long-tail recommendations[9]:** The *TailNet* architecture is designed to enhance long-tail recommendation performance. It classifies items into short-head and long-tail based on click frequency and integrates a unique preference mechanism to balance between recommending niche items for serendipity and maintaining overall accuracy (Liu and Zheng, 2020). The *NISER* method by Gupta et al. (2019) addresses the long-tail issue by focusing on popularity bias in session-based recommendation systems. It aims to ensure item fairness by normalizing item and session representations, thereby improving recommendations, especially for less popular items. Additionally, the above-mentioned approach by Li et al. (2019) also focuses on long-tail recommendations.
- **Self-training mechanisms[10]:** The *Self-Fair* approach by Liu et al. (2022a) employs a self-training mechanism using unlabeled data with the goal of improving user fairness in recommendations for users of different genders. By iteratively refining predictions as pseudo-labels and incorporating fairness constraints, the model balances accuracy and fairness without relying heavily on labeled data.

In the broader context of graph neural networks, researchers have also tackled fairness in non-recommender systems tasks, such as classification (Dai and Wang, 2021; Ma et al., 2021; Dong et al., 2022; Zhang et al., 2022). Their insights provide valuable lessons for future development of fair recommender systems.

7 Discussion and future directions

In this paper, we present a review of the literature on diversity, serendipity/novelty, and fairness in GNN-based recommender systems, with a focus on optimizing for beyond-accuracy metrics. Throughout our analysis, we have explored various aspects of model development and discussed recent advancements in addressing these dimensions.

To further advance the field and guide future research, we have formulated three key questions:

Q1: What are the practical challenges in optimizing GNN-based recommender systems for beyond-accuracy metrics?

GNNs are able to capture complex relationships within graph structures. However, this sophistication can lead to overfitting,

8 Multimodal feature learning integrates diverse data sources, like text, images, and graphs, into unified embeddings to enrich recommendation context and accuracy (Zhou et al., 2023).

9 Long-tail recommendations focus on suggesting less popular or niche items (Kowald et al., 2020).

10 Self-training mechanisms leverage unlabeled data by iteratively predicting and refining labels, enhancing the model's performance with augmented training data. (Yu et al., 2023).

especially when prioritizing accuracy (Fu et al., 2023). Data sparsity and the need for auxiliary data, such as demographic information, challenge the optimization of high-quality node representations, introducing biases (Dhawan et al., 2022). An overemphasis on past preferences can limit novel discoveries (Dhawan et al., 2022), and while addressing popularity bias is essential, it might inadvertently inject noise, reducing accuracy (Liu and Zheng, 2020). Balancing diverse objectives, like fairness, accuracy, and diversity, is nuanced, especially when optimizing one can compromise another (Liu et al., 2022b). These challenges emphasize the need for focused research on effective modeling of GNN-based recommender systems focused on beyond-accuracy optimization.

Q2: Which model development stages of GNN-based recommender systems have seen the most innovation for tackling beyond-accuracy optimization, and which stages have been underutilized?

By conducting a thorough analysis of the reviewed papers (see Table 1), we have observed that the graph construction, propagation layer, and training methodologies have seen significant innovation in GNN-based recommender systems. This includes advanced graph construction methods, innovative graph convolution operations, and unique training methodologies. However, stages like embedding initialization, embedding fusion, and score computation are relatively underutilized. These stages could offer potential avenues for future research and could provide novel ways to balance accuracy, fairness, diversity, and serendipity in recommendations.

Q3: What are potentially unexplored areas of beyond-accuracy optimization in GNN-based recommender systems?

A less explored aspect in GNN-based recommender systems is personalized diversity, which modifies the diversity in recommendations to match individual user preferences. Users favoring more diversity get more diverse recommendations, whereas those liking less diversity get less diverse ones (Eskandanian et al., 2017). This concept of personalized diversity, currently under-researched in GNN-based systems, hints at an intriguing future research direction. It can also relate to personalized serendipity or novelty, tailoring unexpected or novel recommendations to user preferences. Thus, incorporating personalized diversity, serendipity, and novelty in GNN-based systems could enrich beyond-accuracy optimization.

Overall, this review aims to help researchers and practitioners gain a deeper understanding of the multifaceted issues and potential avenues for future research in optimizing GNN-based recommender systems beyond traditional accuracy-centric approaches. By addressing the practical challenges, identifying underutilized model development stages, and highlighting unexplored areas of optimization, we hope to contribute to the development of more robust, diverse, serendipitous, and fair recommender systems that cater to the evolving needs and expectations of users.

Author contributions

TD: literature analysis, conceptualization, and writing. ELa: conceptualization and writing. ELe and DK: conceptualization, writing, and supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the "DDIA" COMET Module within the COMET—Competence Centers for Excellent Technologies Programme, funded by the Austrian Federal Ministry for Transport, Innovation and Technology (bmvit), the Austrian Federal Ministry for Digital and Economic Affairs (bmdw), FFG, SFG, and partners from industry and academia. The COMET Programme is managed by FFG. This research received support by the TU Graz Open Access Publishing Fund. Additional credit is given to OpenAI for the generative AI models, GPT-4, and ChatGPT, used in this work for text summarization and sentence rephrasing. Verification of accuracy and originality was performed for all content generated by these tools.

Conflict of interest

ELa was employed by Infobip.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdollahpouri, H., Adomavicius, G., Burke, R., Guy, I., Jannach, D., Kamishima, T., et al. (2019). Beyond personalization: research directions in multistakeholder recommendation. *arXiv*. doi: 10.48550/arXiv.1905.01986

Abdollahpouri, H., Mansoury, M., Burke, R., Mobasher, B., and Malthouse, E. (2021). "User-centered evaluation of popularity bias in recommender systems," in *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 119–129.

Aggarwal, C. C. (2016). *Recommender Systems, Volume 1*. Cham: Springer.

Boo, S., Kim, S., and Lee, S. (2023). "Serendipity into session-based recommendation: Focusing on unexpectedness, relevance, and usefulness of recommendations," in *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*, 83–86.

Castells, P., Hurley, N., and Vargas, S. (2021). "Novelty and diversity in recommender systems," in *Recommender Systems Handbook*. Cham: Springer, 603–646.

Chen, J., Wu, J., Chen, J., Xin, X., Li, Y., and He, X. (2023). How graph convolutions amplify popularity bias for recommendation? *arXiv*. doi: 10.48550/arXiv.2305.14886

Chen, L., Zhang, G., and Zhou, E. (2018). "Fast greedy map inference for determinantal point process to improve recommendation diversity," in *Advances in Neural Information Processing Systems*, 31.

Chizari, N., Shoeibi, N., and Moreno-García, M. N. (2022). A comparative analysis of bias amplification in graph neural network approaches for recommender systems. *Electronics* 11, 3301. doi: 10.3390/electronics11203301

Dai, E., Zhao, T., Zhu, H., Xu, J., Guo, Z., Liu, H., et al. (2022). A comprehensive survey on trustworthy graph neural networks: privacy, robustness, fairness, and explainability. *arXiv*. doi: 10.48550/arXiv.2204.08570

Dai, E., and Wang, S. (2021). "Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 680–688.

Deldjoo, Y., Jannach, D., Bellogin, A., Difonzo, A., and Zanzonelli, D. (2023). Fairness in recommender systems: research landscape and future directions. *User Modeling and User-Adapted Interaction*, 1–50.

Deldjoo, Y., Noia, T. D., and Merra, F. A. (2021). A survey on adversarial recommender systems: from attack/defense strategies to generative adversarial networks. *ACM Comp. Surv.* (*CSUR)* 54, 1–38. doi: 10.1145/3439729

Dhawan, S., Singh, K., Rabaea, A., and Batra, A. (2022). Improvedgcn: an efficient and accurate recommendation system employing lightweight graph convolutional networks in social media. *Electron. Commer. Res. Appl.* 5, 5:101191. doi: 10.1016/j.elerap.2022.101191

Dong, Y., Liu, N., Jalaian, B., and Li, J. (2022). "Edits: Modeling and mitigating data bias for graph neural networks," in *Proceedings of the ACM Web Conference 2022*, 1259–1269.

Dong, Y., Ma, J., Wang, S., Chen, C., and Li, J. (2023). "Fairness in graph mining: a survey," in *IEEE Transactions on Knowledge and Data Engineering* (IEEE).

Duricic, T., Hussain, H., Lacic, E., Kowald, D., Helic, D., and Lex, E. (2020). "Empirical comparison of graph embeddings for trust-based collaborative filtering," in *Foundations of Intelligent Systems: 25th International Symposium, ISMIS 2020*. Graz, Austria: Springer, 181–191.

Duricic, T., Kowald, D., Schedl, M., and Lex, E. (2021). "My friends also prefer diverse music: homophily and link prediction with user preferences for mainstream, novelty, and diversity in music," in *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 447–454.

Duricic, T., Lacic, E., Kowald, D., and Lex, E. (2018). "Trust-based collaborative filtering: Tackling the cold start problem using regular equivalence," in *Proceedings of the 12th ACM Conference on Recommender Systems*, 446–450.

Ekstrand, M. D., Tian, M., Kazi, M. R. I., Mehrpouyan, H., and Kluver, D. (2018). "Exploring author gender in book rating and recommendation," in *Proceedings of the 12th ACM conference on recommender systems*, 242–250.

Eskandanian, F., Mobasher, B., and Burke, R. (2017). "A clustering approach for personalizing diversity in collaborative recommender systems," in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, 280–284.

Fan, W., Derr, T., Ma, Y., Wang, J., Tang, J., and Li, Q. (2019a). "Deep adversarial social recommendation," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 1351–1357.

Fan, W., Ma, Y., Li, Q., He, Y., Zhao, E., Tang, J., et al. (2019b). "Graph neural networks for social recommendation," in *The World Wide Web Conference*, 417–426.

Fu, Z., Niu, X., and Maher, M. L. (2023). "Deep learning models for serendipity recommendations: A survey and new perspectives," in *ACM Computing Surveys*. New York: Association for Computing Machinery.

Fu, Z., Xian, Y., Gao, R., Zhao, J., Huang, Q., Ge, Y., et al. (2020). "Fairness-aware explainable recommendation over knowledge graphs," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 69–78.

Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., et al. (2023). A survey of graph neural networks for recommender systems: challenges, methods, and directions. *ACM Trans. Recommend. Syst.* 1, 1–51. doi: 10.1145/3568022

Gupta, P., Garg, D., Malhotra, P., Vig, L., and Shroff, G. M. (2019). Niser: normalized item and session representations with graph neural networks. *arXiv*. doi: 10.48550/arXiv.1909.04276

Hamilton, W. L. (2020). Graph representation learning. *Synth. Lect. Artif. Intell. Mach. Learn.* 14, 1–159. doi: 10.1007/978-3-031-01588-5

He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., and Wang, M. (2020). "Lightgcn: Simplifying and powering graph convolution network for recommendation," in *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 639–648.

Isufi, E., Pocchiari, M., and Hanjalic, A. (2021). Accuracy-diversity trade-off in recommender systems via graph convolutions. *Inform. Proc. Manage.* 58, 102459. doi: 10.1016/j.ipm.2020.102459

Kaminskas, M., and Bridge, D. (2016). Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Trans. Interact. Intellig. Syst.* (*TiiS)* 7, 1–42. doi: 10.1145/2926720

Kim, M., Oh, J., Do, J., and Lee, S. (2022). "Debiasing neighbor aggregation for graph neural network in recommender systems," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 4128–4132.

Koren, Y., Rendle, S., and Bell, R. (2021). "Advances in collaborative filtering," in *Recommender Systems Handbook* (Springer), 91–142.

Kotkov, D., Veijalainen, J., and Wang, S. (2020). How does serendipity affect diversity in recommender systems? a serendipity-oriented greedy algorithm. *Computing* 102, 393–411. doi: 10.1007/s00607-018-0687-5

Kowald, D., Schedl, M., and Lex, E. (2020). "The unfairness of popularity bias in music recommendation: a reproducibility study," in *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020* Lisbon, Portugal: Springer, 35–42.

Kunaver, M., and Požrl, T. (2017). Diversity in recommender systems-a survey. *Knowl.-Based Syst.* 123, 154–162. doi: 10.1016/j.knosys.2017.02.009

Lacic, E., Fadljevic, L., Weissenboeck, F., Lindstaedt, S., and Kowald, D. (2022). "What drives readership? an online study on user interface types and popularity bias mitigation in news article recommendations," in *Advances in Information Retrieval: 44th European Conference on IR Research, ECIR 2022*. Stavanger, Norway: Springer, 172–179.

Lacic, E., Kowald, D., Eberhard, L., Trattner, C., Parra, D., and Marinho, L. B. (2015a). "Utilizing online social network and location-based data to recommend products and categories in online marketplaces," in *Mining, Modeling, and Recommending "Things" in Social Media: 4th International Workshops, MUSE 2013*. Paris, France: Springer, 96–115.

Lacic, E., Kowald, D., Traub, M., Luzhnica, G., Simon, J. P., and Lex, E. (2015b). "Tackling cold-start users in recommender systems with indoor positioning systems," in *Proc. of ACM RecSys 15, 2*. New York: ACM.

Leonhardt, J., Anand, A., and Khosla, M. (2018). "User fairness in recommender systems," in *Companion Proceedings of the The Web Conference 2018*, 101–102.

Lex, E., Kowald, D., and Schedl, M. (2020). Modeling popularity and temporal drift of music genre preferences. *Trans. Int. Soc. Music. Inf. Retr.* 3, 17–30. doi: 10.5334/tismir.39

Li, M., Gan, T., Liu, M., Cheng, Z., Yin, J., and Nie, L. (2019). "Long-tail hashtag recommendation for micro-videos with graph convolutional network," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 509–518.

Li, Y., Chen, H., Xu, S., Ge, Y., Tan, J., Liu, S., et al. (2022). Fairness in recommendation: a survey. *arXiv*. doi: 10.48550/arXiv.2205.13619

Liu, H., Lin, H., Fan, W., Ren, Y., Xu, B., Zhang, X., et al. (2022a). Self-supervised learning for fair recommender systems. *Appl. Soft Comput.* 125, 109126. doi: 10.1016/j.asoc.2022.109126

Liu, H., Wang, Y., Lin, H., Xu, B., and Zhao, N. (2022b). Mitigating sensitive data exposure with adversarial learning for fairness recommendation systems. *Neural Comp. Appl.* 34, 18097–18111. doi: 10.1007/s00521-022-07373-4

Liu, S., Ounis, I., Macdonald, C., and Meng, Z. (2020). "A heterogeneous graph neural model for cold-start recommendation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2029–2032.

Liu, S., and Zheng, Y. (2020). "Long-tail session-based recommendation," in *Proceedings of the 14th ACM Conference on Recommender Systems*. New York: ACM, 509–514.

Liu, Z., Ma, Y., Ouyang, Y., and Xiong, Z. (2021). Contrastive learning for recommender system. *arXiv*. doi: 10.48550/arXiv.2101.01317

Ma, J., Deng, J., and Mei, Q. (2021). Subgroup generalization and fairness of graph neural networks. *Adv. Neural Inf. Process. Syst.* 34, 1048–1061.

Ma, X., Hu, Q., Gao, Z., and AbdelHady, M. (2022). "Contrastive co-training for diversified recommendation," in *2022 International Joint Conference on Neural Networks (IJCNN)*. Padua: IEEE, 1–9.

Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., and Burke, R. (2020). "Fairmatch: A graph-based approach for improving aggregate diversity in recommender systems," in *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, New York: ACM, 154–162.

Pu, P., Chen, L., and Hu, R. (2012). Evaluating recommender systems from the user's perspective: survey of the state of the art. *User Model. User-adapt. Interact.* 22, 317–355. doi: 10.1007/s11257-011-9115-7

Rendle, S., Freudenthaler, C., Gantner, Z., and Schmidt-Thieme, L. (2012). "BPR: Bayesian personalized ranking from implicit feedback," in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, 452–461.

Silveira, T., Zhang, M., Lin, X., Liu, Y., and Ma, S. (2019). How good your recommender system is? A survey on evaluations in recommendation. *Int. J. Mach. Learn. Cybern.* 10, 813–831. doi: 10.1007/s13042-017-0762-9

Skarding, J., Gabrys, B., and Musial, K. (2021). Foundations and modeling of dynamic networks using dynamic graph neural networks: a survey. *IEEE Access* 9, 79143–79168. doi: 10.1109/ACCESS.2021.3082932

Sun, J., Guo, W., Zhang, D., Zhang, Y., Regol, F., Hu, Y., et al. (2020). "A framework for recommending accurate and diverse items using bayesian graph convolutional neural networks," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery* & *Data Mining*, 2030–2039. doi: 10.1145/3394486.3403254

Wang, C., Liang, Y., Liu, Z., Zhang, T., and Philip, S. Y. (2021). "Pre-training graph neural network for cross domain recommendation" in *2021 IEEE Third International Conference on Cognitive Machine Intelligence (CogMI)* (IEEE), 140–145.

Wang, H., Zhang, F., Zhang, M., Leskovec, J., Zhao, M., Li, W., et al. (2019). "Knowledge-aware graph neural networks with label smoothness regularization for recommender systems" in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery* & *Data Mining*, 968–977.

Wang, X., He, X., Wang, M., Feng, F., and Chua, T.-S. (2019). "Neural graph collaborative filtering," in *Proceedings of the 42nd international ACM SIGIR Conference on Research and development in Information Retrieval*, 165–174. doi: 10.1145/3331184.3331267

Wang, Y., Ma, W., Zhang, M., Liu, Y., and Ma, S. (2023a). A survey on the fairness of recommender systems. *ACM Trans. Inform. Syst.* 41, 1–43. doi: 10.1145/3547333

Wang, Y., Zhao, Y., Zhang, Y., and Derr, T. (2023b). Collaboration-aware graph convolutional network for recommender systems. *Proc. ACM Web Conf.* 2023, 91–101. doi: 10.1145/3543507.3583229

Wu, K., Erickson, J., Wang, W. H., and Ning, Y. (2022). "Equipping recommender systems with individual fairness via second-order proximity embedding," in *2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. Istanbul: IEEE, 171–175.

Wu, S., Sun, F., Zhang, W., Xie, X., and Cui, B. (2022). Graph neural networks in recommender systems: a survey. *ACM Comp. Surv.* 55, 1–37. doi: 10.1145/3535101

Xie, R., Liu, Q., Liu, S., Zhang, Z., Cui, P., Zhang, B., et al. (2021). Improving accuracy and diversity in matching of recommendation with diversified preference network. *IEEE Trans. Big Data* 8, 955–967. doi: 10.1109/TBDATA.2021.3103263

Xu, C., Zhang, Y., Chen, H., Dong, L., and Wang, W. (2023). A fairness-aware graph contrastive learning recommender framework for social tagging systems. *Inf. Sci.* 640, 119064. doi: 10.1016/j.ins.2023.119064

Yang, L., Wang, S., Tao, Y., Sun, J., Liu, X., Yu, P. S., et al. (2023). "Dgrec: Graph neural network for recommendation with diversified embedding generation," in *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 661–669.

Yang, Y., Huang, C., Xia, L., Huang, C., Luo, D., and Lin, K. (2023). "Debiased contrastive learning for sequential recommendation," in *Proceedings of the ACM Web Conference 2023*. New York: ACM, 1063–1073.

Ye, R., Hou, Y., Lei, T., Zhang, Y., Zhang, Q., Guo, J., et al. (2021). "Dynamic graph construction for improving diversity of recommendation," in *Proceedings of the 15th ACM Conference on Recommender Systems*, 651–655.

Yin, H., Cui, B., Li, J., Yao, J., and Chen, C. (2012). "Challenging the long tail recommendation," in *Proceedings of the VLDB Endowment* (VLDB Endowment), 896–907.

Yu, J., Yin, H., Xia, X., Chen, T., Li, J., and Huang, Z. (2023). "Self-supervised learning for recommender systems: a survey," in *IEEE Transactions on Knowledge and Data Engineering* (IEEE).

Zhao, Y., Wang, Y., Liu, Y., Cheng, X., Aggarwal, C., and Derr, T. (2023). Fairness and diversity in recommender systems: a survey. *arXiv preprint arXiv:2307.04644*. doi: 10.48550/arXiv.2307.04644

Zhang, H., Wu, B., Yuan, X., Pan, S., Tong, H., and Pei, J. (2022). Trustworthy graph neural networks: Aspects, methods and trends. *arXiv*. doi: 10.48550/arXiv.2205.07424

Zhao, M., Wu, L., Liang, Y., Chen, L., Zhang, J., Deng, Q., et al. (2022). "Investigating accuracy-novelty performance for graph-based collaborative filtering," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 50–59.

Zheng, Y., Gao, C., Chen, L., Jin, D., and Li, Y. (2021). "Dgcn: Diversified recommendation with graph convolutional networks," in *Proceedings of the Web Conference* 2021, 401–412. doi: 10.1145/3442381.3449835

Zhou, H., Zhou, X., Zeng, Z., Zhang, L., and Shen, Z. (2023). A comprehensive survey on multimodal recommender systems: taxonomy, evaluation, and future directions. *arXiv*.

Zhou, T., Kuscsik, Z., Liu, J.-G., Medo, M., Wakeling, J. R., and Zhang, Y.-C. (2010). Solving the apparent diversity-accuracy dilemma of recommender systems. *Proc. Nat. Acad. Sci.* 107, 4511–4515. doi: 10.1073/pnas.1000488107

Zhou, Y., Zheng, H., Huang, X., Hao, S., Li, D., and Zhao, J. (2022). Graph neural networks: Taxonomy, advances, and trends. *ACM Trans. Intel. Syst. Technol. (TIST)* 13, 1–54. doi: 10.1145/3495161

Zuo, Y., Liu, S., and Zhou, Y. (2023). Dtgcf: Diversified tag-aware recommendation with graph collaborative filtering. *Appl. Sci.* 13, 2945. doi: 10.3390/app13052945

frontiers | Frontiers in Big Data

TYPE Review
PUBLISHED 26 February 2024
DOI 10.3389/fdata.2024.1304439

Check for updates

OPEN ACCESS

EDITED BY
Emanuel Lacić,
Infobip d.o.o., Croatia

REVIEWED BY
Bettina Fazzinga,
University of Calabria, Italy
Simone Kopeinik,
Know Center, Austria

*CORRESPONDENCE
Mathias Uta
mathias.uta@siemens-energy.com

RECEIVED 29 September 2023
ACCEPTED 06 February 2024
PUBLISHED 26 February 2024

CITATION
Uta M, Felfernig A, Le V-M, Tran TNT, Garber D, Lubos S and Burgstaller T (2024) Knowledge-based recommender systems: overview and research directions.
Front. Big Data 7:1304439.
doi: 10.3389/fdata.2024.1304439

COPYRIGHT
© 2024 Uta, Felfernig, Le, Tran, Garber, Lubos and Burgstaller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Knowledge-based recommender systems: overview and research directions

Mathias Uta[1]*, Alexander Felfernig[2], Viet-Man Le[2], Thi Ngoc Trang Tran[2], Damian Garber[2], Sebastian Lubos[2] and Tamim Burgstaller[2]

[1]Siemens Energy AG, Erlangen, Germany, [2]Institute of Software Technology (IST) - Applied Software Engineering & Ai Research Group (ASE), Graz University of Technology, Graz, Austria

Recommender systems are decision support systems that help users to identify items of relevance from a potentially large set of alternatives. In contrast to the mainstream recommendation approaches of collaborative filtering and content-based filtering, knowledge-based recommenders exploit semantic user preference knowledge, item knowledge, and recommendation knowledge, to identify user-relevant items which is of specific relevance when dealing with complex and high-involvement items. Such recommenders are primarily applied in scenarios where users specify (and revise) their preferences, and related recommendations are determined on the basis of constraints or attribute-level similarity metrics. In this article, we provide an overview of the existing state-of-the-art in knowledge-based recommender systems. Different related recommendation techniques are explained on the basis of a working example from the domain of survey software services. On the basis of our analysis, we outline different directions for future research.

KEYWORDS

recommender systems, semantic recommender systems, knowledge-based recommender systems, case-based recommendation, constraint-based recommendation, critiquing-based recommendation, constraint solving, model-based diagnosis

1 Introduction

Recommender systems support users in identifying relevant items from a large set of alternatives and thus help to reduce the complexity of decisions and increase user satisfaction and sales (Felfernig and Burke, 2008; Burke et al., 2011; Knijnenburg et al., 2012). *Single-shot recommender systems* recommend items based on already stored preferences of a user (e.g., videos a user purchased/liked in the past) without the need of engaging the user in a dialog for capturing further preferences (Rafter and Smyth, 2005; Pramod and Bafna, 2022). In contrast, *conversational recommender systems* perform preference elicitation on the basis of a user/recommender dialog (Goeker and Thompson, 2000; Bridge, 2002; Gao et al., 2021). Depending on the application scenario, different recommendation approaches can be applied (see Tables 1, 2 for an overview of used data sources and properties of those recommendation approaches).

First, collaborative filtering (CF) (Ekstrand et al., 2011) is based on the idea of simulating word-of-mouth promotion by determining recommendations on the basis of the preferences of so-called nearest neighbors (NNs), which are users with preferences similar to those of the current user. For example, a movie that has been watched by the nearest neighbors of the current user (but not by the current user) and has been evaluated positively by those nearest neighbors should also be recommended to the current user. A major advantage of CF approaches is an easy setup since no detailed information about the provided items is needed. Furthermore, serendipity effects can be created, i.e., collaborative filtering is a good choice for finding recommendations which are (in a positive sense) a kind of "surprise" for the user, thus supporting diversity and the identification of completely unexpected/unsearched items (Iaquinta et al., 2008; Ziarani and Ravanmehr, 2021). CF typically follows the idea of single-shot recommendation where no session-specific user dialogs are needed. A disadvantage of CF recommenders is cold-start problems which are due to the sparsity of rating information (regarding users and items) available in the recommendation algorithm, making it hard to identify accurate recommendations (Lika et al., 2014).

Second, content-based filtering (CBF) (Pazzani and Billsus, 2007) is based on the idea of recommending items which are in one way or another similar to items the user has consumed in the past. In contrast to CF, CBF does not exploit the preferences of nearest neighbors which limits the cold-start problem to the availability of initial user item ratings. A kind of disadvantage of CBF recommendation approaches is need of item knowledge, for example, in terms of item categories and/or item descriptions which are used to derive a user profile representing a user item preference history. CBF systems are typically of type "single-shot", i.e., no conversational user interface needs to be provided for preference elicitation purposes. *Basic CBF* that purely focuses on the recommendation of similar items is less appropriate when it comes to the triggering of serendipity effects (Iaquinta et al., 2008). For example, if a user has purchased the song *Personal Jesus* from *Depeche Mode*, an item that could be recommended in the future could be the same song performed by *Johnny Cash*. However, further developments in content-based recommendation have improved the capability of CBF with regard to the achievement of serendipity effects, for example, on the basis of increasing the diversity of recommended items (Maccatrozzo et al., 2017). In the line with CF approaches, CBF has an easy setup process, since only item knowledge is needed in the setup phase. Since preferences change over time, both CF and CBF have an issue with item domains characterized by infrequently articulated and changing preferences. For example, time intervals between individual car purchases of a user could be around 10 years—in such long time periods, user preferences could completely change, which makes the identification of recommendations challenging without using a preference elicitation dialog.

Third, knowledge-based recommender (KBR) systems can be considered as complementary to CF- and CBF-based approaches in terms of avoiding the related cold-start difficulties (Burke, 2000; Towle and Quinn, 2000; Lorenzi and Ricci, 2005). KBR systems are based on the idea of collecting the preferences of a user (preference elicitation) within the scope of a dialog and then to recommend items either (1) on the basis of a predefined set of recommendation rules (constraints) or (2) using similarity metrics that help to identify items which are similar to the preferences of the user. The first approach is denoted as *constraint-based recommendation* (Felfernig and Burke, 2008), whereas the second one is referred to as *case-based recommendation* (Lorenzi and Ricci, 2005)—these two can be regarded as major types of knowledge-based recommender systems (Aggarwal, 2016). Serendipity effects in knowledge-based recommendation are limited by the static encoding in terms of constraints (rules) and similarity metrics. KBR systems support the determination of recommendations specifically in complex and high-involvement item domains [domains where suboptimal decisions can have significant negative consequences, for example, when investing in high-risk financial services (Felfernig et al., 2006)] where items are not bought on a regular basis (Aggarwal, 2016). Example item domains are financial services (Felfernig et al., 2007; Musto et al., 2015), software services (Felfernig et al., 2021), apartment or house purchasing (Fano and Kurth, 2003), and digital cameras (Felfernig et al., 2006). These systems are able to take into account constraints (e.g., high-risk financial services must not be recommended to users with a low preparedness to take risks) and provide explanations of recommendations also in situations where no solution could be identified. In contrast to CF and CBF, KBR systems support explicit preference elicitation dialogs which makes them immune with regard to user preferences changing over time. Due to an often time-intensive knowledge exchange between domain experts and knowledge engineers, the definition of recommendation knowledge can trigger high setup costs (Ulz et al., 2017). Both types of KBR systems (case-based and constraint-based) are conversational, since a preference elicitation dialog helps to figure out user preferences (Gao et al., 2021).

1.1 Further recommendation approaches

In addition to the three basic approaches of collaborative filtering, content-based filtering, and knowledge-based recommendation, there exist various further approaches. First, *hybrid recommender systems* (Burke, 2002) exploit synergy effects by combining different recommendation approaches. For example, by combining CBF with CF, the cold-start problem can be solved by initially applying CBF in order to collect the relevant rating data needed for establishing a CF-based approach. Furthermore, *group recommender systems* (Masthoff, 2015; Dara et al., 2020; Felfernig et al., 2024b) support the determination of recommendations for whole groups, i.e., not individual users. These systems apply basic recommendation approaches in such a way that recommendations satisfy the preferences of all or at least a subset of the group members. For example, the *average* user rating per item can be regarded as an item rating provided by the whole group.

1.2 Article contributions

The major contributions of this article are as follows. First, we provide an overview of the existing state-of-the-art in knowledge-based recommender systems. Our work significantly enhances existing overviews on the same topic (see the last row of Table 3) specifically in terms of the inclusion of new technological approaches and a broader view on how knowledge-based technologies are applied to recommender systems. Second, to assure understandability, we explain different underlying recommendation techniques on the basis of a working example. Third, to foster further related work, we discuss different open research directions derived from our literature analysis.

TABLE 1 Basic recommendation approaches of collaborative filtering (CF), content-based filtering (CBF), and knowledge-based recommendation (KBR), and used data sources including the current user preferences, a user's preference history, the preferences of nearest neighbors, and item knowledge.

Approach	Preferences	Preference history	Nearest neighbors	Item knowledge
Collaborative (CF)	–	×	×	–
Content-based (CBF)	–	×	–	×
Knowledge-based (KBR)	×	–	–	×

TABLE 2 *Basic properties* of different recommendation approaches: *easy setup* = low effort needed for setting up the recommender system, *dialog-based* = conversational process between system and user, *serendipity* = effect of proposing unexpected but relevant recommendations, *cold-start problem* = initial data are needed to provide reasonable recommendations, *high-involvement items* = a user carefully evaluates the candidate items since suboptimal decisions can have significant negative consequences.

Recommendation approach	Collaborative (CF)	Content-based (CBF)	Knowledge-based (KBR)
Easy setup	×	×	–
Dialog-based	–	–	×
Serendipity	×	–	×
Cold-start problem	×	×	–
High-involvement items	–	–	×

The remainder of this article is organized as follows. In Section 2, we explain the methodology applied for our literature analysis. An introduction to basic knowledge-based recommendation approaches is presented in Section 3, where we specifically focus on techniques and applications of case-based and constraint-based recommendation. Thereafter, in Section 4, we provide an overview of advanced techniques in knowledge-based recommendation with a specific focus on integration scenarios with machine learning. In this context, we also discuss related topics such as hybrid recommendation with knowledge-based recommenders and knowledge-based group recommender systems. In Section 5, we discuss the identified research directions. The article is concluded with Section 6.

2 Methodology

Our overview of the existing state-of-the-art in knowledge-based recommendation (KBR) is based on a literature analysis conducted using the activities of *search* (querying of leading research portals), *review* (evaluating and classifying the identified scientific contributions), and *discussion of reviewed contributions* (deciding about inclusion of the identified studies in our overview based on the selection criteria of quality and relevance). Queries have been performed between December 2023 and January 2024 on the research platforms Google Scholar,[1] ResearchGate,[2] ScienceDirect,[3] SpringerLink,[4] and Elsevier,[5] using the keywords of "recommender systems" + ["knowledge-based" | "case-based" | "critiquing" | "ontologies" | "knowledge graphs" | "constraint satisfaction" | "conversational" | "mass customization" | "groups" | "overview"], representing 10 individual search queries. Using these queries, we have also analyzed topic-related scientific contribution in conferences and journals: the International Joint Conference on Artificial Intelligence (IJCAI), the AAAI Conference on Artificial Intelligence, the European Conference on Artificial Intelligence (ECAI), the Recommender Systems Conference (RecSys), the User Modeling, Adaptation, and Personalization (UMAP) conference, the Constraint Programming Conference (CP), the Software Product Line Conference (SPLC), User Modeling and User-Adapted Interaction (UMUAI), and ACM Transactions on Recommender Systems. Thereafter, using the snowballing technique (Wohlin, 2014), we have analyzed the reference sections of the identified studies. Major criteria for estimating the relevance of a paper were (1) a *topic-wise match*, i.e., a relationship to knowledge-based recommenders and (2) an *official publication* in a workshop, conference, journal, magazine, book, or PhD thesis. With our analysis, we have identified 97 publications directly related to the topic of knowledge-based recommendation that have been used as a basis for this article (see Table 3).

3 Basic approaches and applications

In this section, we introduce a working example from the domain of *survey software service recommendation* that is used to explain different knowledge-based recommendation approaches. The underlying scenario is that users search for a survey software configuration that fulfills their preferences, for example, a researcher received a new project funding and is now interested in purchasing a survey software that supports different user studies envisioned for the new project. In many cases, Web applications (services) are preferred over solutions requiring an installation at the customer site. In such a scenario, knowledge-based recommenders can support users in identifying an appropriate

1 https://scholar.google.com/
2 https://www.researchgate.net/
3 https://www.sciencedirect.com/
4 https://link.springer.com/
5 https://www.elsevier.com/

TABLE 3 Overview of scientific contributions in knowledge-based recommender systems organized in the line of the primary focus of the paper: ALG, algorithms; PREF, preference elicitation; KA, knowledge acquisition; APP, application; FUR, further recommendation approaches and knowledge representations; OV, overview articles/papers including aspects related to knowledge-based recommendation.

Approach	ALG	PREF	KA	APP
Case-based (CBR)	Jannach, 2006	Goeker and Thompson, 2000; Bridge, 2002; McSherry, 2003; Mirzadeh et al., 2005; Christakopoulou et al., 2016	Khan and Hoffmann, 2003; Zou et al., 2020	Fesenmaier et al., 2003; Lee and Kim, 2015; Musto et al., 2015; Feely et al., 2020; Bokolo, 2021; Hernandez-Nieves et al., 2021
Critiquing-based (CRIT)	Reilly et al., 2004; Smyth et al., 2004; McCarthy et al., 2005; Mandl and Felfernig, 2012; Murti et al., 2016	Zhang et al., 2008; Chen et al., 2017; Xie et al., 2018; Wu et al., 2019; Güell et al., 2020	McCarthy et al., 2010	Burke et al., 1996; Grasch et al., 2013
Constraint-based (CON)	Cöster et al., 2002; Felfernig and Burke, 2008; Falkner et al., 2011; Fargier et al., 2016; Erdeniz et al., 2019; Teppan and Zanker, 2020; Felfernig et al., 2023a	Towle and Quinn, 2000; Fano and Kurth, 2003; Junker, 2004; Felfernig et al., 2009b, 2012, 2013c, 2018a,b; Pereira et al., 2016; Tazl et al., 2019; Erdeniz et al., 2022; Uta et al., 2022	Jannach and Kreutler, 2007; Felfernig et al., 2009a, 2013a,b, 2015; Daoudi et al., 2016; Uta et al., 2021; Lubos et al., 2023	Felfernig and Kiener, 2005; Felfernig et al., 2006, 2007; Zanker et al., 2006; Murphy et al., 2015; Wobcke et al., 2015; Ulz et al., 2017; Almalis et al., 2018
Further (FUR)	McCarthy et al., 2006	Burke, 2002; Zhou et al., 2020; Zhu et al., 2020; Le et al., 2022	Wang et al., 2019a,b; Sun et al., 2020; Esheiba et al., 2021; Sha et al., 2021	Lee et al., 2006; Bahramian and Ali Abbaspour, 2015; Colombo-Mendoza et al., 2015; Pessemier et al., 2017; Cordero et al., 2020; Dong et al., 2020
Overviews (OV)	Bridge et al., 2005; Felfernig et al., 2011, 2024b; Aggarwal, 2016; Gao et al., 2021	Lorenzi and Ricci, 2005; Chen and Pu, 2012; Masthoff, 2015; Atas et al., 2021; Felfernig et al., 2021; Jannach et al., 2021; Tran et al., 2023	Felfernig, 2007; Bouraga et al., 2014; Felfernig et al., 2014; Cena et al., 2021; Popescu et al., 2022	Burke, 2000; Felfernig et al., 2024a

configuration (including pricing) of a survey software service that fulfills their wishes and needs. In our example, we include the user-selectable survey software features *ABtesting* (should ABtesting be supported by the survey software), *statistics* (should a basic statistics feature be included), *multiplechoice* (should the specification of multiple-choice questions be possible), and *license* [which license model is preferred—*free of charge* (0) vs. *with costs*, i.e., 100]. In the following, we show how knowledge-based recommenders can support the identification of relevant items, i.e., survey software configurations.

Knowledge-based recommender systems can operate on different knowledge representations—these knowledge representations will be explained and exemplified in Section 3.1. Thereafter, we introduce the two basic approaches to knowledge-based recommendation, which are (1) *case-based recommendation* including *critiquing-based recommendation* as a specific form of case-based recommendation (see Section 3.2) and (2) *constraint-based recommendation* (see Section 3.3).

3.1 Recommendation knowledge representations

Knowledge representations of knowledge-based recommender systems can be (1) *table-based* which is used in scenarios where items are represented in terms of product table entries or (2) *constraint-based* which is used in scenarios where items are defined on the basis of a set of *restrictions* (also denoted as *rules* or *constraints*). In the first case (extensional representation—see, for example, Table 4), each item that could be recommended is explicitly defined in a corresponding item (product) table. In the second case, there is no need to enumerate all items since items are specified in a constraint-based fashion (intensional representation—see, for example, Table 5).

3.1.1 Table-based representations

In many knowledge-based recommendation scenarios, the offered itemset is represented in terms of an item (product) table (see Table 4). The itemset is defined extensionally, i.e., all selectable alternatives (items) are enumerated. In our example, five different service configurations (items $i_1..i_5$) can be offered to a user. Table-based representations can be applied if the set of offered items is limited, i.e., the item space is rather small which is often the case, for example, in digital camera or financial service recommendation (Felfernig et al., 2006, 2007). Using a table-based knowledge representation, corresponding database queries can be performed to identify a set of recommendation candidates that support the preferences defined by the user (Felfernig et al., 2006, 2023a). An example of a *user preference* regarding the itemset defined in Table 4 could be $ABtesting = 1$, meaning that the user is interested in a survey software that includes (supports) *ABtesting*. For simplicity, we assume that attributes have a corresponding Boolean domain definition, for example, the domain of attribute *ABtesting* is $\{0, 1\}$ where $1 = true$ (feature included) and $0 = false$. One exception

TABLE 4 Extensional itemset representation where each item is represented as a table entry assuming the item (table) attributes *ABtesting* (0,1), *statistics* (0,1), *multiplechoice* (0,1), and *license* (0,100).

Item	ABtesting	Statistics	Multiplechoice	License
i_1	0	1	0	0
i_2	0	0	1	100
i_3	1	1	0	100
i_4	0	1	1	100
i_5	1	1	1	100

In our working example, the offered survey software configurations (items $i_1..i_5$) represent the complete set of items (i.e., the product/item catalog). For example, item i_1 represents a survey software configuration (a "free of charge version") which does not support *ABtesting* and *multiplechoice* questions, i.e., only single choice questions can be asked and the *statistics* feature supports a basic analysis of survey data.

TABLE 5 Intensional solution space representation in terms of a set of constraints $\{c_1..c_5\}$ on the CSP variables *ABtesting*, *statistics*, *license*, and *multiplechoice*.

Id	Constraint
c_1	ABtesting=1 → statistics=1
c_2	¬(ABtesting=1 ∧ license=0)
c_3	¬(multiplechoice=1 ∧ license=0)
c_4	¬ multiplechoice=1 ∧ ¬ ABtesting=1 → license=0
c_5	¬(ABtesting=0 ∧ statistics=0 ∧ multiplechoice=0 ∧ license=0)

thereof is the *license* attribute with a domain $\{0, 100\}$ representing the price of a license.

3.1.2 Constraint-based representations

Alternatively, itemsets can be represented in an intensional fashion on the basis of a set of domain-specific constraints (see Table 5), for example, as a *constraint satisfaction problem* (CSP) (Rossi et al., 2006; Felfernig and Burke, 2008). Such knowledge representations are specifically useful if the solution (item) space becomes intractable, i.e., defining and maintaining all alternatives is extremely inefficient and error-prone (or even impossible) and related search queries become inefficient and at least impractical for interactive settings (Falkner et al., 2011).

The constraints in Table 5 represent exactly the solution space, as shown in Table 4. If such constraints are not explicitly known, a given set of items can be defined in terms of one constraint in disjunctive normal form, for example, the entries in Table 4 would be represented as $(ABtesting = 0 \wedge statistics = 1 \wedge license = 0 \wedge multiplechoice = 0) \vee .. \vee (ABtesting = 1 \wedge statistics = 1 \wedge license = 100 \wedge multiplechoice = 1)$—for details, we refer to Felfernig and Burke (2008).

3.1.3 Further knowledge representations

Intensional representations of solution spaces (itemsets) as presented in Table 5 can be implemented with different knowledge representations ranging from *constraint satisfaction problems* (CSPs) (Rossi et al., 2006; Felfernig and Burke, 2008) and *Boolean satisfiability problems* (SAT problems) (Biere et al., 2021; Felfernig et al., 2021) to less frequently used recommendation knowledge representations such as *answer set programming* (ASP) (Eiter et al., 2009; Teppan and Zanker, 2020) and *ontology-based knowledge representations*, for example, *description logics* (DL) (Lee et al., 2006; McGuinness, 2007). In addition, database queries can be applied in such a way that intensionally formulated recommendation knowledge is encoded directly in database queries—see, for example, Felfernig et al. (2006, 2023a). Without loss of generality, in this article, we focus on CSP-based representations when discussing the concepts of constraint-based recommendation. As a basis for our discussion of case-based recommendation approaches, we will use the entries of Table 4.

3.2 Case-based recommendation

3.2.1 Basic approach

Following the basic concepts of case-based reasoning (Kolodner, 2014), case-based recommendation uses a knowledge-rich representation of the item domain and can therefore be classified as a kind of knowledge-based recommendation (Burke, 2000; Khan and Hoffmann, 2003; Lorenzi and Ricci, 2005). In a basic setting, the item assortment can be regarded as the case base, and recommendations are determined by identifying those items from the case base which "support" the user preferences (Lorenzi and Ricci, 2005). In this context, product features (also denoted as properties or attributes) are used to specify user preferences. The more preferences are supported by an item, the higher its user relevance. Examples of case-based recommendation approaches going beyond an equality match between item properties and user preferences are the following—see, for example, Lorenzi and Ricci (2005).

3.2.2 Interest confidence value

This approach is based on the idea of predicting item relevance on the basis of the similarity of a new item with items a user has already evaluated positively in the past (Lorenzi and Ricci, 2005; Musto et al., 2015). This approach is in the line with the ideas of content-based recommendation where the similarity between a new item and properties of already consumed items is used to estimate recommendation relevance. A simple example of the

TABLE 6 Example of an *interest confidence value* (*icv*)-based evaluation of the user relevance of new items (in_k) on the basis of the similarity with already consumed items i_l.

Consumed items	New items				
	in_1	in_2	in_3	in_4	in_5
i_1	0.2	0.4	0.3	0.9	0.5
i_2	0.4	0.6	0.2	0.6	0.7
i_3	0.4	0.1	0.5	0.5	0.6
icv (avg)	0.33	0.36	0.33	0.66	0.6

For example, the similarity between the consumed item i_1 and the new item in_2 is assumed to be 0.4. In this example, item in_4 has the highest *interest confidence value* (0.66) and therefore can be regarded as a user-relevant item.

application of interest confidence values (*icv*) is shown in Table 6, where the *average* (*avg*) *similarity* between a new item (the user did not see up to now) and (consumed) items positively evaluated in the past is used to estimate item relevance.

3.2.3 Attribute-level similarity metrics

This approach is based on the idea that the degree of satisfaction of individual user requirements has a direct impact on the corresponding item ranking (Lorenzi and Ricci, 2005). In such contexts, attribute-level similarity metrics (McSherry, 2003) can be used to determine the similarity between the user requirements and the items included in the product assortment. For example, if explicit price information of an item is available, the *less is better* (LIB) similarity metric can be applied (the lower the price, the higher the item ranking). The *more is better* (MIB) similarity metric can be applied in the context of technical attributes, such as the resolution of a digital camera or the return rate of a financial service (McSherry, 2003; Felfernig et al., 2013c) (the higher the attribute value the better). *Equal is better* (EIB) is used in contexts where users explicitly require a specific attribute value (e.g., the color of a car), and *nearer is better* (NIB) can be used in situations where an item should fulfill a specific requirement as good as possible, for example, the size of a TV screen in the living room. The overall similarity between user requirements and an item is then determined on the basis of a "global" similarity that combines individual attribute-level similarity functions. In our working example, the *license* attribute could be associated with an *LIB*, the remaining attributes with an *EIB* attribute-level similarity.

3.2.4 Refine, relax, and compromise

When querying a case base (i.e., an item catalog) with a set of user preferences, it can happen that the number of candidate items (items that satisfy the user preferences) is (1) too large and—as a consequence—the initial set of user requirements needs to be refined or (2) too small (in the worst case, no solution could be identified), which means that the user requirements have to be relaxed. In the first case, the set of user requirements needs to be extended, for example, if a user has specified the requirement $statistics = 1$, this results in four candidate items ($\{i_1, i_3, i_4, i_5\}$—see Table 4). Refining the original query to $statistics = 1 \wedge ABtesting = 1$ reduces the set of candidate items to $\{i_3, i_5\}$. In contrary, if a user has specified the requirement $ABtesting = 1 \wedge license = 0$, the corresponding query would result in an empty set. The user has two options to resolve the inconsistency, i.e., to relax the query: either to exclude *ABtesting* or to accept a *license* payment, i.e., $license = 100$. Such trade-off decisions can be supported by so-called *compromise-based user interfaces* which group items with regard to different possible compromises (McSherry, 2003). Such interfaces would group and explain candidate items in the line of, for example, *most of your requirements are fulfilled, however, these items require a license payment*.

3.2.5 Critiquing

Critiquing-based recommendation originates from different case-based recommenders (Kolodner, 1992; Bridge et al., 2005) which often support a search-based approach where—depending on a set of defined user preferences—the system recommends items which support in one way or another those preferences. While basic case-based and constraint-based recommendation focus on supporting a *search-based recommendation process*, critiquing-based recommender systems (as a specific type of case-based recommender system) follow a *navigation-based approach* (Chen and Pu, 2012), which focuses on better supporting users in *exploring* the solution (item) space. Critiquing-based recommendation is based on the idea of presenting *example (reference) items* to a user who then can (1) accept the proposed item or (2) define critiques in terms of changes that are needed to make an item acceptable. Existing critiquing-based recommenders are based on a predefined itemset (product table)—see Table 4. These systems are regarded as knowledge-based, since items are associated with semantic properties, and user-defined (selected) critiques can be regarded as logical criteria (constraints) to be fulfilled by recommendations.

3.2.6 Unit critiquing

Let us assume that in the context of our working example a critiquing-based recommender supports six basic critiques which are (1) *furtherstat*, (2) *reducestat*, (3) *excludemc*, (4) *includemc*, (5) *excludelicense*, and (6) *includelicense*. The critiquing-based approach used in this example is *unit-critiquing* (Chen and Pu, 2012; Mandl and Felfernig, 2012), where in each critiquing cycle, a critique on an individual item attribute is specified. Table 7 shows how critiques can be translated into corresponding item selection criteria. For example, if a user selects the critique *further statistics*

TABLE 7 Translating critiques into item selection criteria where *mc=multiplechoice*, *stat=statistics*, and "↦" represents a mapping operator from critiques to item selection criteria.

Critique	Item selection criteria
Furtherstat	$Statistics = 0 \mapsto statistics = 1$
	$Statistics = 1 \mapsto ABtesting = 1$
Reducestat	$ABtesting = 1 \mapsto ABtesting = 0$
	$Statistics = 1 \wedge ABtesting = 0 \mapsto statistics = 0$
Excludemc	$Multiplechoice = 1 \mapsto multiplechoice = 0$
Includemc	$Multiplechoice = 0 \mapsto multiplechoice = 1$
Excludelicense	$License = 100 \mapsto license = 0$
Includelicense	$License = 0 \mapsto license = 100$

For example, if the critique *excludelicense* is specified by the user, an item supporting the selection criteria $license = 0$ will be shown.

(*furtherstat*) and statistics is not included in the shown reference item, the corresponding selection criteria would be $statistics = 1$. Furthermore, if statistics is already included in the reference item, $ABtesting = 1$ would be defined as additional selection criteria (see the entries in Table 7).

Table 8 shows an example of a *unit critiquing session* where the initial (reference) item presented to the user is i_1. The user specifies the critique *furtherstat* which requires the inclusion of additional statistic features into the survey software service configuration. Taking into account, this critique means to combine the properties of the reference item i_1 ($ABtesting = 0, statistics = 1, license = 0, multiplechoice = 0$) with the selection criteria derived from Table 7 (i.e., $ABtesting = 1$), resulting in a new set of criteria ($\{ABtesting = 1, statistics = 1, license = 0, multiplechoice = 0\}$). There does not exist an item in $\{i_1..i_5\}$ which completely fulfills these criteria. A related tradeoff is to choose an item as a new reference item which is mostly similar to the current selection criteria—in our case, this is item i_3 which only requires the adaptation of $license = 0$ to $license = 100$. The user again specifies a critique (*includemc*) which results in a new set of criteria ($\{ABtesting = 1, statistics = 1, license = 100, multiplechoice = 1\}$), leading to the new reference item i_5 which is finally accepted by the user.

The idea of critiquing-based recommendation is to identify an item which is similar to the current item but (in addition) takes into account the criteria specified by the critique. In our example, we have assumed that the search for a new item returns an exact match (*EIB* metric), i.e., all search criteria are fulfilled. However, the search for new reference items is often designed more flexible, for example, other attribute-level similarity metrics can be applied to determine new reference items (Lorenzi and Ricci, 2005).

3.2.7 Further critiquing approaches

For demonstration purposes, we have discussed the concepts of unit critiquing in more detail. However, critiquing-based recommender systems support various alternative types of critiques (Chen and Pu, 2012; Güell et al., 2020). The idea of *compound critiques* (Smyth et al., 2004; Zhang et al., 2008) is to allow the definition of critiques which refer to more than one item property. A related example would be the compound critique *furtherstat and includelicense*, indicating the interest of a user in further statistics features also accepting (additional) license costs. A major advantage of compound critiquing is that due to the specification of multiple criteria in one critiquing cycle, significantly larger "jumps" in the itemspace are possible (Aggarwal, 2016) and—as a result—less critiquing cycles are needed to find a solution. As an extension of compound critiques, *dynamic critiques* (Reilly et al., 2004; McCarthy et al., 2005) are based on the idea of mining frequent critiquing patterns which are then presented to a user.

3.2.8 Handling inconsistent search criteria

Critiquing-based recommender systems try to take into account the critiques defined by a user. Since user preferences (search criteria) typically change within the scope of a recommendation session, inconsistencies can occur (Atas et al., 2021). For example, a user is first interested in a survey software configuration without license fee resulting in the reference item i_1. Then, the user wants to additionally include the *ABtesting* feature which then results in an inconsistency, since no item supports both, no license fee and *ABtesting* at the same time. From the logical point of view, such (inconsistent) search criteria are constraint sets $CRIT = \{crit_1..crit_q\}$ and in the case of inconsistencies in $CRIT$, one option is to determine explanations (diagnoses) which indicate different options of resolving an inconsistency.[6] Alternatively, weighting schemes are applied in such contexts meaning that "elder" critiques have a higher probability of being deleted from a critiquing history. Another alternative is the retrieval of a new reference item which is as much as possible similar to the current set of search criteria (see our example in Table 8).

3.2.9 Examples of case-based recommender systems

Table 9 provides an overview of example case-based recommender applications. (Burke et al., 1996) introduce the FindMe approach to support a unit critiquing-based navigation in complex information spaces—discussed item domains are cars, videos, and apartments. The Entree system (Burke, 2000) supports unit critiquing-based recommendations in the restaurant domain, and this system is also based on the FindMe approach introduced by (Burke et al., 1996). DieToRecs (Fesenmaier et al., 2003) is a case-based recommender system that supports a similarity-based approach to case retrieval, i.e., users have to specify their requirements and select a case of relevance which is then used as a basis for further search. Recomment (Grasch et al., 2013) is a natural language-based user interface with an underlying unit-critiquing-based recommender system. (Musto et al., 2015) introduce an asset allocation strategy framework which supports the recommendation of asset classes (e.g., Euro

6 For details, see the following discussion of constraint-based recommender systems.

TABLE 8 A simple example of a unit-critiquing session consisting of two critiquing cycles.

Attribute	Ref. item (i_1)	Ref. item after cycle 1 (i_3)	Ref. item after cycle 2 (i_5)
ABtesting	0	1	1
Statistics	1	1	1
License	0	100	100
Multiplechoice	0	0	1
Unit critique	(1) further statistics	(2) includemc	Accept

In critiquing cycle (1), the critique *further statistics* is specified on the reference item i_1, in cycle (2), *includemc* is specified as a critique. After cycle (2), the user is satisfied with (accepts) the recommended item (i_5).

TABLE 9 Overview of example case-based recommender applications.

Recommender application	Recommendation domain	References
RENTME	Apartments	Burke et al., 1996
ENTREE	Restaurants	Burke, 2000
DIETORECS	Travel plans	Fesenmaier et al., 2003
RECOMMENT	Digital cameras	Grasch et al., 2013
Asset allocation framework	Financial services	Musto et al., 2015
EHEARSS	Healthcare	Lee and Kim, 2015
Recommender for Recreational Runners	Training plans	Feely et al., 2020
CEBRA	Financial services	Hernandez-Nieves et al., 2021
Smart City Initiative	Smart cities	Bokolo, 2021

Bond, High Yield Bond, Euro Stocks, and Emerging Market Stocks) that fulfill the goals of the investor and are basically generated on the basis of an analysis of the portfolios of similar clients (investors). Lee and Kim (2015) introduce EHEARSS which is a case-based recommender system in the healthcare domain which helps to identify, for example, relevant doctors, depending on the current (aggregated) health record and those of similar cases. In this context, an ontology helps to assure the correctness of recommendations, for example, in terms of the proposed medical diagnosis. Feely et al. (2020) introduce a case-based recommendation approach for the recommendation of marathon training plans and race pacings based on case information about the workouts and race histories of similar runners. CEBRA (Hernandez-Nieves et al., 2021) supports the recommendation of banking products using a basic case-based reasoning recommender operating on a user's financial service profile and corresponding demographic data. Finally, the smart city initiative (Bokolo, 2021) focuses on the case-based recommendation of smart city dimensions, i.e., by taking a look at similar cities, corresponding smart city-related activities for the current city are determined in order to achieve predefined sustainability goals, for example, on the basis of the best use of technological and human resources.

3.3 Constraint-based recommendation

The concept of constraint-based recommendation (Felfernig and Burke, 2008) is based on the idea that recommendation knowledge is represented in terms of a set of variables and a corresponding set of constraints. As already mentioned, such constraints can be defined in an extensional fashion (by just enumerating the items part of the solution space) or in an intensional fashion where the recommendation space is represented in terms of a set of logical formulae. In this context, a constraint-based recommendation task (CB-REC TASK) can be defined as follows (see Definition 1).[7]

Definition 1. A constraint-based recommendation task (CB-REC TASK) can be defined as a constraint satisfaction problem (CSP) (V, C, R) where $V = \{v_1..v_n\}$ is a set of finite domain variables with associated variable domain definitions $dom(v_i)$, $C = \{c_1..c_m\}$ is a set of constraints, and $R = \{r_1..r_k\}$ is a set of user requirements.

Following Definition 1, an example recommendation task (V, C, R) based on the entries in Table 5 is as follows: $V = \{ABtesting, statistics, multiplechoice, license\}$, $dom(ABtesting) = dom(statistics) = dom(multiplechoice) = \{0, 1\}, dom(license) = \{0, 100\}$, $C = \{c_1..c_5\}$, and $R = \{r_1 : ABtesting = 1\}$, assuming that the user is interested in the *ABtesting* feature.

A constraint-based recommendation CB-REC for a given constraint-based recommendation task (CB-REC TASK) can be defined as follows (see Definition 2).

Definition 2. A constraint-based recommendation (CB-REC) for a defined CB-REC TASK is a set of tuples $REC = \bigcup_{i_\alpha \in I}\{(rank_{i_\alpha}, i_\alpha)\}$ where $rank_{i_\alpha}$ represents the recommendation rank assigned to item $i_\alpha \in I$ defined by a recommendation function rf and $\forall(rank_{i_\alpha}, i_\alpha) \in REC$: consistent($C \cup R \cup a(i_\alpha)$) where $a(i_\alpha)$ denotes the variable value assignments associated with item i_α.

Assuming $R = \{r_1 : ABtesting = 1\}$, a CB-REC in our working example could be $REC = \{(1, i_5), (2, i_3)\}$ where $a(i_3) = \{ABtesting = 1, statistics = 1, license = 100, multiplechoice = 0\}$ and $a(i_5) = \{ABtesting = 1, statistics = 1, license = 100, multiplechoice = 1\}$.

7 For simplicity, we omit a differentiation between variables describing user preferences and those describing item properties and different constraints between these variables—for related details, we refer to Felfernig et al. (2006) and Felfernig and Burke (2008).

3.3.1 Ranking items

Up to now, we did not specify a function rf for the ranking of the items in REC. A simple ranking function could just count the number of supported features (see Equation 1), which would result in the mentioned recommendation ranking of $REC = \{(1, i_5), (2, i_3)\}$ since the number of supported features of i_5 is 3 [$support(i_5) = 3$] whereas $support(i_3) = 2$. In other words, item i_3 is outperformed by one item resulting in a ranking of 2.[8]

$$rf(i_\alpha) = 1 + |\{i_k \in REC(k \neq \alpha) : support(i_k) > support(i_\alpha)\}| \quad (1)$$

An alternative to our simplified ranking function (Equation 1) is to introduce a utility function which evaluates utility of individual items on the basis of a pre-defined set of interest dimensions (Felfernig et al., 2006, 2018a). In our software service recommendation scenario, examples of relevant interest dimensions are *economy* and *quality* (see also Table 10).

Such utility-based evaluation schemes can then be used by a utility function to determine the overall utility of individual items—see, for example, Equation (2)— where D represents a set of interest dimensions [in our case, $D = \{economy, quality\}$] and $eval(i_\alpha, d)$ represents the evaluation scheme as presented in Table 10. Assuming equal importance of the two example interest dimensions [e.g., $importance(economy) = 0.5$ and $importance(quality) = 0.5$], $utility(i_3) = 0.0 + 10.0 = 10.0$ and $utility(i_5) = 0.0 + 15 = 15.0$, i.e., item i_5 has a higher utility compared with item i_3. Depending on the user-specific importance of individual interest dimensions, the resulting utility values can differ.

$$utility(i_\alpha) = \Sigma_{d \in D}\, eval(i_\alpha, d) \times importance(d) \quad (2)$$

3.3.2 Dealing with inconsistent requirements

In constraint-based recommendation, it can be the case that individual user requirements do not allow the determination of a recommendation.[9] For example, if a user is interested in the *ABtesting* and *multiplechoice* features but does not want to pay a license, i.e., $R = \{ABtesting = 1, multiplechoice = 1, license = 0\}$, no solution/item can support this set of requirements. In this example, we are able to identify two different conflicts (Junker, 2004) (see Definition 3), which are minimal sets of requirements that induce an inconsistency.

Definition 3. A conflict (set) $CS \subseteq R$ is a set of constraints with inconsistent($CS \cup C$), i.e., no solution can be found for $CS \cup C$. A conflict set CS is minimal if $\neg\exists CS' : CS' \subset CS$ (subset minimality).

Conflict set minimality is important due to the fact that just one requirement needs to be deleted (i.e., relaxed) from CS in order to resolve the conflict. In our example, there exist two minimal conflict sets which are CS_1:$\{ABtesting = 1, license = 0\}$ and CS_2:$\{multiplechoice = 1, license = 0\}$. To resolve the conflict CS_1, a user can choose between the two options of excluding *ABtesting* ($ABtesting = 0$) or paying a *license* ($license = 100$). The same approach can be followed to resolve CS_2, i.e., to either accept $multiplechoice = 0$ or accept $license = 100$.

Such conflict sets can be used in interactive recommendation settings to support users in figuring out ways from the *no solution could be found* dilemma. A set Δ of requirements is denoted as diagnosis (Reiter, 1987; Felfernig et al., 2009b, 2012) if it helps to resolve all identified conflicts (see Definition 4).

Definition 4. A diagnosis $\Delta \subseteq R$ is a set of constraints with consistent($R - \Delta \cup C$), i.e., at least one solution can be found for $R - \Delta \cup C$. A diagnosis Δ is minimal if $\neg\exists\Delta' : \Delta' \subset \Delta$ (subset minimality).

Both concepts, i.e., *conflict sets* and corresponding *diagnoses* can be used to support users in inconsistent situations. Conflict sets are helpful in the context of *repeated conflict resolution* (for example, when purchasing a new car for the family, the preferences of individual family members change over time), and diagnoses can be used when users are interested in *quick repairs* (for example, when parametrizing an operating system, some inconsistent parameter settings need to be adapted).

3.3.3 Examples of constraint-based recommender systems

Table 11 provides an overview of example constraint-based recommender applications. FSADVISOR (Felfernig and Kiener, 2005) is a financial service recommender system that proposes new financial services depending on the current financial status and requirements of the customer. In the line of FSADVISOR, VITA (Felfernig et al., 2007) is a financial service recommender system supporting sales representatives in the preparation and conduction of customer dialogs. Both systems are based on a database query-based approach implemented in ADVISORSUITE, which is a development environment for constraint-based recommender applications (Felfernig et al., 2006; Jannach and Kreutler, 2007). In addition, MORTIMER is a constraint-based recommender application developed on the basis of ADVISORSUITE which focuses on the constraint-based recommendation of cigars. AUTHENTIC (Murphy et al., 2015) is a constraint satisfaction-based recommendation environment for supporting the achievement of energy saving goals on the basis of adapted schedules for activating household appliances. Recommendations of behavior change (e.g., postponed activation of a dishwasher to exploit more self-produced energy) are determined on the basis of energy consumption-related sensor data. Wobcke et al. (2015) introduce a P2P recommendation environment in the context of online dating scenarios where constraints (rules) are used to define baseline recommendation strategies. RECTURK (Ulz et al., 2017) is a framework for the development of constraint-based recommender applications. The underlying idea is to apply the concepts of *human computation* (Law and von Ahn, 2011) to alleviate the definition and maintenance of recommendation rules. Almalis et al. (2018) introduce a constraint-based job recommender system where constraints are used to assure defined compatibilities between job seekers and job announcements. Finally, Esheiba et al. (2021)

8 In this context, *license* is not regarded as a feature.

9 In critiquing-based recommendation, such situations occur if search criteria become inconsistent.

TABLE 10 A simple utility-based evaluation scheme for recommendation candidates, for example, paying no license fee, i.e., *license* = 0, contributes to the evaluation dimension economy.

Interest dimension	ABtesting		Statistics		License		Multiplechoice	
	0	1	0	1	0	100	0	1
Economy	0	0	0	0	10	0	0	0
Quality	0	10	0	10	0	0	0	10

TABLE 11 Overview of example constraint-based recommender applications.

Recommender application	Recommendation domain	References
FSAdvisor	Financial services	Felfernig and Kiener, 2005
Advisor Suite	Digital cameras	Felfernig et al., 2006
Mortimer	Cigars	Zanker et al., 2006
VITA	Financial services	Felfernig et al., 2007
Authentic	Energy saving	Murphy et al., 2015
P2P framework	Online dating	Wobcke et al., 2015
RecTurk framework	Digital cameras	Ulz et al., 2017
FoDRA framework	Jobs	Almalis et al., 2018
PSS recommender	Product services	Esheiba et al., 2021

present a constraint-based recommender system supporting the filtering-out of irrelevant product services and the ranking of the remaining relevant ones.

4 Recent advances in knowledge-based recommendation

In this section, we provide an overview of the recent advances in knowledge-based recommendation (compared with the basic approaches presented in Section 3). We discuss the following aspects: (1) Users are often in the situation of not knowing every detail about a product and—as a consequence—are not able to make a related informed decision. For example, if someone is interested in digital cameras, he/she might be able to specify preferences regarding attributes such as *resolution*, *maximum price*, *weight*, and *primary application* (e.g., landscape photography or sports) but at the same time might not be able to specify preferences regarding attributes such as *maximum frames per second* or *maximum exposure time*. In our example, users might not be aware of the usefulness of the *ABtesting* feature. Related techniques to support users in such contexts are presented in Section 4.1. (2) Given a set of user preferences, situations can occur where no solution can be identified (see also Section 3)—in Section 4.2, we introduce concepts for personalized conflict resolution. (3) In some situations, users have adaptation (reconfiguration) requirements regarding already purchased items. For example, after having purchased an online survey software license, a user detected the relevance of the *ABtesting* feature and is interested in a feature extension. Related technical solutions are presented in Section 4.3. (4) Knowledge-based recommendation is not only relevant in recommendation scenarios focusing on single users but also in scenarios where recommendations have to be determined for groups, for example, planning round-trips for tourist groups or deciding about the features to be included in a software or service. Aspects of knowledge-based group recommender systems are presented in Section 4.4. (5) In Section 4.5, we discuss further approaches focusing on the topics of hybrid recommendation, search optimization, explanations, recommendation knowledge acquisition, conversational recommendation, and explanations.

4.1 Recommending preference settings

There are situations where users need help in specifying preferences regarding specific item properties. Table 12 depicts a simplified example of an attribute value proposal in the context of a *case-based recommendation* scenario: the *current* user has already specified his/her preferences regarding the attributes *multiplechoice* and *statistics*. Using a nearest neighbor (NN)-based approach, i.e., determining the session with the most similar preferences, results in session s_3. Consequently, we can recommend the attribute values *ABtesting* = 1 and *license* = 100 to the *current* user. The same approach to attribute value recommendation can also be applied (1) in the context of customer segmentation (recommendation) where customers have to be identified who will be interested in a new item feature (Felfernig et al., 2021) and (2) to figure out features (attributes), a customer is interested in specifying—this is relevant in scenarios where users are not interested in specifying all attributes (Mirzadeh et al., 2005; Dragone et al., 2018).

An alternative to NN-based recommendation is to apply model-based approaches such as matrix factorization or neural networks—for related details see Erdeniz et al. (2019, 2022) and Uta et al. (2022).

Beyond case-based recommendation, such approaches to attribute value recommendation are also applied in *constraint-based recommendation* (and beyond) (Felfernig and Burke, 2008; Falkner et al., 2011; Fargier et al., 2016; Pereira et al., 2016; Temple et al., 2017). A major issue in such contexts is the fact that a recommendation of one or more attribute settings could be *inconsistent* with the underlying set of constraints (C). One way to assure consistency is to pre-determine a set of nearest neighbors and to show recommendations to users after having tested recommendation consistency (Cöster et al., 2002; Felfernig and Burke, 2008; Falkner et al., 2011; Pereira et al., 2016).

TABLE 12 Example user sessions s_i representing "successfully" completed recommendation sessions (e.g., a user has selected a specific item).

Session	ABtesting	Statistics	Multiplechoice	License
s_1	1	1	0	100
s_2	0	1	0	0
s_3 (NN)	1	1	1	100
Current	?	1	1	?

In this example, user (session) s_3 can be regarded as the nearest neighbor (NN) of the *current* user, i.e., both, *ABtesting* and *license* are recommended to be included.

An alternative is to learn *solver search heuristics* in such a way that constraint solvers are enabled to predict user-relevant attribute value settings (Uta et al., 2022), i.e., to directly integrate recommendation knowledge into constraint solver search, for example, by applying neural networks that receive as input a set of user preferences and propose corresponding solver search heuristics [in terms of variable (value) orderings] as output. In the example shown in Table 12, variable value orderings for the not yet instantiated variables *ABtesting* and *license* would be *ABtesting* [1,0] and *license* [100,0], indicating that the solver should first try to instantiate these variables with 1 (100) before trying other instantiations. For an overview on the integration of machine learning and constraint solving, we refer to Popescu et al. (2022).

In *critiquing-based recommendation*, such attribute value recommendations could be also used in the initial preference specification phase. However, critiquing-based recommendation follows the idea of critiquing a shown reference item. Specifically, in critiquing scenarios, attribute value recommendation is "replaced" by the recommendation of critiques (McCarthy et al., 2010; Mandl and Felfernig, 2012; Murti et al., 2016; Xie et al., 2018)—see the simplified unit critique recommendation example depicted in Table 13. In this simplified example,[10] the critiquing session most similar to the critiquing session of the *current* user (session) is cs_2. This can be determined, for example, by determining the intersection of the critiquing sets of cs_i and *current* where $|critiques(cs_1) \cap critiques(current)| = 1$ and $|critiques(cs_2) \cap critiques(current)| = 2$. Using a nearest neighbor-based approach, critiques and item selection of cs_2 can be used to recommend future critiques to the *current user* but also to predict items that will be chosen by the current user (Mandl and Felfernig, 2012).

4.2 Dealing with "no solution could be found" situations

As discussed in Section 3, users of knowledge-based recommenders in some situations need support to get out of the *no solution could be found dilemma*. Approaches that can proactively support users in such contexts are conflict detection (Junker, 2004) and model-based diagnosis (Reiter, 1987; Felfernig et al., 2012; Walter et al., 2017). Such algorithms help users to understand trade-offs in the current situation and propose different options to resolve inconsistencies. Having user preference weights available, for example, in terms of explicitly defined preference weights, algorithms can be applied in different ways to determine preferred (personalized) diagnoses. For demonstration purposes, we assume the following preference ordering (see Table 14) for our example conflict sets $CS1: \{ABtesting = 1, license = 0\}$ and $CS2$: $\{multiplechoice = 1, license = 0\}$ derived from the user requirements $R = \{ABtesting = 1, multiplechoice = 1, license = 0\}$.

Following a simple additive utility-based scheme (Felfernig et al., 2013c) using the orderings depicted in Table 14 (the lower the ordering position, the higher the importance of the related user requirement), the conflicts CS_1 and CS_2 would be resolved (in a personalized fashion) as follows: for the user in Session s_1 (user 1), we would keep the exclusion of *license* as-is and exclude both, *ABtesting* and *multiplechoice* ($1 < 3 + 4$). Vice-versa, in the case of session s_2 (user 2), we would keep the preferences regarding *ABtesting* and *multiplechoice* as-is and accept the inclusion of a license fee ($1 + 2 < 4$). In a similar fashion, such an approach can be applied in critiquing-based recommendation sessions where—on the basis of importance information about individual critiques [user sentiments about individual attributes could be available or "older" critiques could be interpreted as less important (Chen et al., 2017)]—corresponding conflict resolutions could be proposed.

4.3 Recommendations for reconfiguration

In scenarios where users have already purchased an item and want to extend the item afterward, recommendations for reconfigurations are required, i.e., which attributes of the original solution need to be adapted such that the new user requirements can be taken into account (Felfernig et al., 2018b; Weckesser et al., 2018). For example, if a user has already ordered a specific survey software service consisting of the feature $\{statistics = 1\}$ (item i_1) but then detects that he/she wants to include also the *ABtesting* feature, the question arises which of the initial attribute settings (user preferences) remain the same and which ones need to be adapted. In such scenarios, critiquing-based recommendation would not be the primary choice since reconfiguration is in the need of *minimal changes* to an already ordered item. In contrast, case-based reasoning can be applied by using cases (items) with an exact attribute-wise match to the already purchased item with a minimal number of needed adaptations.

If we assume the availability of a product table as shown in Table 4 and a user who has purchased item i_1 and now wants to extend this configuration with *ABtesting*, this results in the search criteria $\{ABtesting = 1, statistics = 1, multiplechoice = 0, license =$

10 For simplicity, all critiquing sessions have the same number of critiques, however, in real-world scenarios, the number of critiques can very significantly between different critiquing sessions.

TABLE 13 Example critiquing sessions cs_i representing "successfully" completed recommendation sessions.

Session	Reference item	$crit_1$	$crit_2$	$crit_3$	Selection
cs_1	i_1	i_1.*furtherstat*	i_4.*reducestat*	-	i_1
cs_2 (NN)	i_1	i_1.*includemc*	i_4.*furtherstat*	i_5.*excludemc*	i_3
Current	i_1	i_1.*includemc*	i_7.*furtherstat*	?	?

In this example, session cs_2 can be regarded as the nearest neighbor (NN) of the current user. The reference item i_1 is presented in Table 4.

TABLE 14 Example preference orderings for requirements (the lower the more important the preference).

Session	ABtesting	Statistics	Multiplechoice	License	Diagnosis Δ
s_1	3	2	4	1	$\{ABtesting = 1, multiplechoice = 1\}$
s_2	2	3	1	4	$\{license = 0\}$

For example, in session s_1, keeping the selected *license* type as-is has the highest priority.

TABLE 15 Simplified reconfiguration setting: a user has already purchased item i_1 and—after some time—wants to extend this configuration by including *ABtesting*.

Item	ABtesting	Statistics	Multiplechoice	License
i_1 (purchase)	0 (1)	1	0	0
i_2	0	0	1	100
i_3 (NN)	1	1	0	100
i_4	0	1	1	100
i_5	1	1	1	100

0}. The corresponding nearest neighbor configuration is item i_3 [based on the *equal is better* (EIB) similarity metric]—see Table 15. In addition to the inclusion of *ABtesting*, the selection of item i_3 requires one additional adaptation which is $license = 100$. Beyond counting the number of adaptations, reconfiguration can also take into account sentiments (Chen et al., 2017): if additional features could be included without further costs or efforts, this should be preferred over reconfigurations where those features are not included.

In constraint-based recommendation, $i = \{a_1..a_k\}$ can be regarded as a set of variable assignments (constraints) describing the original item i (in our example, i_1); furthermore, $R = \{r_1..r_l\}$ is a set of reconfiguration requirements (in our example, $R = \{ABtesting = 1\}$). A diagnosis $\Delta \subseteq i$ is a set of variable assignments in i that need to be adapted such that a solution for the reconfiguration task can be identified (Felfernig et al., 2018b). More formally, $\Delta \subseteq i$ represents a diagnosis if consistent ($i - \Delta \cup C \cup R$), i.e., the corresponding domain-specific constraints C (see Table 5) need to be taken into account (Felfernig et al., 2018b). If weights for user preferences are available (see, e.g., Table 14), reconfiguration can be personalized, i.e., for individual users having purchased the same original item, different reconfigurations could be proposed. Following the standard approach of conflict detection (Junker, 2004) and diagnosis identification (Reiter, 1987), a reconfiguration can be determined the same way as shown in the context of identifying minimal sets of inconsistent requirements. If a user has originally selected (purchased) item i_1, a corresponding reconfiguration requirement $R = \{ABtesting = 1\}$ would induce the (singleton) conflicts CS_1: $\{ABtesting = 0\}$ and CS_2: $\{license = 0\}$, resulting in a corresponding reconfiguration $\{ABtesting = 1, statistics = 1, multiplechoice = 0, license = 100\}$ (Felfernig et al., 2018b).

4.4 Knowledge-based recommendation for groups

Group recommender systems (Pessemier et al., 2017; Felfernig et al., 2024b) are typically built on the top of single user recommender systems following one of the approaches of (1) *aggregated predictions* or (2) *aggregated models*. With aggregated predictions, the items recommended to individual group members can be merged. In the context of aggregated models, the preferences of individual group members are merged, resulting in a group preference model which is then used to determine recommendations.

In *case-based recommendation* for groups, the determination of group recommendations can be based on two different approaches. With *aggregated predictions*, in a first step, relevant recommendations would be determined for each user (see Table 16). In the following, these individual recommendations (items) have to be aggregated, according to a specific aggregation strategy. For example, if we apply *majority voting*, the overall recommendation for the group would be item i_5 since it has received more recommendations compared with other recommendation candidates.

If the recommender system is based on *aggregated models* (see Table 17), user-individual preferences have to be aggregated into a group model which can then be used for determining

TABLE 16 Example of user-individual item recommendations as a basis for determining a group recommendation using the *aggregated prediction* approach (with majority voting, i_5 gets recommended).

User	Item
u_1	i_5
u_2	i_4
u_3	i_1
u_4	i_5

recommendations. In the example of Table 17, the overall case-based group recommendation would be item i_3 since this one exactly represents the preferences of the determined group model.

We want to emphasize that following the approach of *aggregated models* can result in situations where no exact match between the attribute settings in the group model and corresponding item properties is possible. However, in such cases, the applied similarity functions will recommend items most similar to the preferences stored in the group model.

If a *critiquing-based recommendation* approach (McCarthy et al., 2006) is used, an *aggregated prediction* approach can be implemented in such a way that each group member completes an individual critiquing process, and the resulting items are then merged by an aggregation function. Following an *aggregated models* approach, individual critiques can be merged, for example, after one or more critiquing cycles, in order to show a new reference item (to individual group members) which is regarded as current recommendation for the whole group—see also McCarthy et al. (2006) and Felfernig et al. (2024b).

In *constraint-based recommendation*, in both cases, i.e., aggregated preferences and aggregated models, a group recommender would have to aggregate the preferences (attribute settings) of the individual group members (Felfernig et al., 2024b). Again, different aggregation approaches on the attribute level are possible—if we follow the goal of achieving fairness (Le et al., 2022), for example, in terms of a nearly equal number of accepted compromises per group member, a corresponding optimization (minimization) problem needs to be solved (see Equation 3). In this context, $p(u_\alpha)$ denotes the preferences of group member u_α, *rec* denotes the determined group recommendation, and $diff(p(u_\alpha, rec))$ denotes the number of preference tradeoffs to be accepted by group member u_α. This is a group recommendation that takes into account the aspect of *fairness* minimizing tradeoff distances among individual group members.

$$MIN \leftarrow \Sigma_{\{u_i,u_j\}\subseteq Users}\, |diff\ (p(u_i, rec) - diff\ (p(u_j, rec))| \qquad (3)$$

4.5 Further aspects

4.5.1 Hybrid recommendation approaches

The knowledge-based recommendation approaches discussed up to now have—as a central element—an item assortment (i.e., product catalog) or a knowledge base that describes the item assortment in an intensional fashion. Knowledge-based recommendation can also be applied in hybrid recommendation scenarios taking over the role of a supportive/complementary component (Burke, 2002). In many cases, knowledge-based recommender systems take over the role of being responsible of checking domain-specific properties, for example, when recommending medical services, it must be assured that the medical item is compatible with the current state of health of a patient. Furthermore, knowledge bases (e.g., in terms of ontologies and knowledge graphs) can be applied, for example, as a basis for similarity metrics that are able to take into account knowledge structures (Bahramian and Ali Abbaspour, 2015; Colombo-Mendoza et al., 2015; Wang et al., 2019a,b; Sun et al., 2020; Zhu et al., 2020; Sha et al., 2021).

4.5.2 Search optimization

Performance optimization can become an issue specifically in the context of constraint-based recommendation and related conflict detection and diagnosis problem solving (Reiter, 1987; Junker, 2004; Felfernig et al., 2012; Le et al., 2023). Improving the efficiency of constraint solving primarily depends on the ability to select optimal heuristics depending on the given search problem. Optimizing search efficiency is often based on the application of machine learning. A detailed overview of integration scenarios for constraint solving and machine learning is given in the study by Popescu et al. (2022). Specifically in the context of constraint-based recommendation, multi-criteria optimization becomes an issue since (1) constraint-based recommenders are typically applied in interactive scenarios with the need of efficient solution search and (2) at the same time, solutions must be personalized, i.e., take into account the preferences of a user. Thus, constraint-based recommendation is an important application field with the need of an in-depth integration of knowledge-based reasoning and machine learning (Uta et al., 2022).

4.5.3 Knowledge acquisition and maintenance

This is an important aspect specifically in the context of constraint-based recommendation where recommendation knowledge is encoded in terms of constraints, rules, or database queries (Jannach and Kreutler, 2007; Felfernig et al., 2009a, 2013b). It must be ensured that the defined constraints correspond with real-world constraints, i.e., reflect the item domain and recommendation knowledge. The related phenomenon of the *knowledge acquisition bottleneck*, i.e., significant communication overheads between domain experts and knowledge engineers, is still omnipresent when applying such knowledge-based systems. Related research efforts focus on the topics of *automated testing and debugging*, *human-centered knowledge acquisition*, and *deep models of knowledge understanding*. Automated testing and debugging focuses on the application of model-based diagnosis where pre-defined test cases are used to induce conflicts in the knowledge base which are then resolved on the basis of diagnosis algorithms—see, for example, Le et al. (2021). Understanding the complexity individual knowledge structures can also help to increase the maintainability of knowledge bases—this topic includes the aspects of cognitive complexities of knowledge structures (Felfernig et al., 2015) and related knowledge structuring, for example, in terms

TABLE 17 Example of a group recommendation based on the *aggregated model* approach resulting in the recommendation of item i_3.

Session	ABtesting	Statistics	Multiplechoice	License
u_1	1	1	0	100
u_2	0	1	0	0
u_3	1	1	1	100
Group model	1	1	0	100

of the ordering of constraints in a recommendation knowledge base (Felfernig et al., 2013a). Finally, human-computation-based knowledge acquisition (Ulz et al., 2017) and user query recommendation (Daoudi et al., 2016) make knowledge definition accessible to domain experts by asking simple questions, for example, *given the user requirements of statistics* = 1 *and license* = 100*, which items can be recommended?* From the answers to such questions, a recommendation knowledge base can be generated (Ulz et al., 2017).

4.5.4 Conversational recommendation

Following the characterizations of Christakopoulou et al. (2016), Wu et al. (2019), Cordero et al. (2020), Dong et al. (2020), Zhou et al. (2020), and Jannach et al. (2021) conversational recommender systems are interactive systems supporting users in the navigation of the item space in an efficient fashion and recommend items based on a systematic approach of preference elicitation. Knowledge-based recommenders can be regarded as conversational since users are guided through a preference elicitation dialog (Zou et al., 2020). However, further types of conversational recommender systems exist which extend existing approaches specifically with more flexible types of preference elicitation using, for example, natural language dialogs (Grasch et al., 2013), chatbot technologies (Tazl et al., 2019), and specific interfaces based on large language models (Dai et al., 2023).

4.5.5 Explanations in knowledge-based recommendation

In the context of recommender systems, explanations can have various goals such as persuading users to purchase an item, to increase the level of trust in a recommendation, or simply to extend a user's item domain knowledge (Tintarev and Masthoff, 2012). In this context, basic explanation types are the following: (1) *why* explanations help a user to understand the reasons why a specific item has been recommended. In the context of single-shot recommendation approaches such as collaborative filtering and content-based filtering, explanations are directly related to the used algorithmic approach. For example, *this item is recommended, since similar users also liked it* or *this item is recommended, since you liked similar items in the past*. Answering such *why* questions in the context of knowledge-based recommendation means to relate recommendations to user preferences, for example, *this camera is recommended since it includes a high frame rate per second which corresponds to your requirement to be able to perform sports photography on a professional level*. This example also shows that recommendation dialogs do not necessarily include technical item properties but could also support scenarios where users specify *high-level preferences* (Felfernig et al., 2006) which are then translated into technical properties by the recommender system. (2) *Why not* explanations help users to understand in more detail why no solution could be found for their requirements. In this context, conflicts (Junker, 2004) help to understand, for example, individual incompatibilities between user requirements, whereas diagnoses (Reiter, 1987; Felfernig et al., 2012) are a general proposal (explanation) for resolving an inconsistency. (3) *How* explanations are more related to specific aspects of a recommendation process, for example, when using a utility-based approach for item ranking (Felfernig et al., 2006), explanations can take into account corresponding weights to explain how a recommendation has been determined, for example, *since you have ranked the priority of the feature fps (frame rate per second) very high, item X is the one which is ranked highest since it received the highest utility value on the basis of our evaluation function. How* explanations play a major role in the context of group recommendations—specifically to achieve goals such as fairness and consensus (Tran et al., 2023; Felfernig et al., 2024b). An example of taking into account such aspects is the following: *since the preferences of user X have not been taken into account for already 3 months, the current restaurant recommendation specifically focuses on taking into account the preferences of X.*

5 Research directions

Basic insights from our analysis are the following. A major strength of knowledge-based (specifically constraint-based) recommenders is their ability to enforce domain-specific constraints. These systems are useful when recommending and explaining high-involvement items such as cars and financial services. A major disadvantage is "setup" efforts that are needed to predefine the recommendation knowledge in terms of product properties, product catalogs, similarity metrics, and constraints. There are many new developments in *knowledge-based (recommender) systems* focusing on a "deep" integration with *machine learning*, thus helping to unify the two worlds (Popescu et al., 2022). However, there are still a couple of research directions that are of interest—see the following discussion.

5.1 Integrating knowledge-based systems with machine learning

Specifically in constraint-based recommender systems, a major line of research focuses on the integration of concepts and techniques from machine learning into constraint-based reasoning, for example, by learning variable and variable value ordering heuristics (Popescu et al., 2022; Uta et al., 2022). This way, recommendation functionalities can be directly integrated with constraint solving. An issue for future research is to analyze alternative integrations, for example, in which way constraint-based reasoning can be integrated into machine learning—such a goal can be achieved using neuro-symbolic AI concepts (Monroe, 2022).

5.2 Diagnosis performance optimization

A major runtime performance bottleneck in knowledge-based recommenders are conflict detection and diagnosis algorithms which operate on the basis of individual consistency checks (Jannach, 2006). A related topic for future research is to intensively integrate machine learning techniques that can help to reduce the number of irrelevant consistency checks as much as possible and thus help to improve runtime performance of conflict detection and diagnosis.

5.3 Cognitive aspects of preference elicitation

Human decision making is amenable to different types of biases with a potential negative impact on decision quality (Atas et al., 2021). Recommender user interfaces and corresponding preference elicitation dialogs have to take this aspect into account. A major issue for future research is to understand in detail in which way recommender user interfaces can have an impact on decision processes and how to counteract decision biases in such contexts.

5.4 Evaluation metrics for knowledge-based recommenders

Current approaches to evaluate recommender systems primarily focus on the aspect of prediction quality and user acceptance (Uta et al., 2021). However, recommender systems also have a huge impact on how customers perceive items, which items are selected, and—as a consequence—which items have to be produced. This way, recommender systems have to be evaluated with regard to various additional aspects such as the visibility of items in recommendation lists (e.g., how often is an item highly-ranked) and restrictiveness of specific features in recommendation dialogs (e.g., are there too few items that support specific feature combinations). Initial related work on these aspects can be found in the study by Lubos et al. (2023).

5.5 Consequence-based explanations

Explanations in recommender systems focus on reasons as to why specific items have been recommended, why no solution could be identified, and how items have been determined (Friedrich and Zanker, 2011; Tintarev and Masthoff, 2012). Future research should focus more on the generation of explanations that take into account the consequences of choosing specific items, for example, in the context of investment decisions, a consequence-based explanation could be of type *since you prefer to purchase a quite expensive car, this will have an impact on both, the quality of living (in terms of having a smaller house) and in terms of being able to have money for holidays*.

TABLE 18 Examples of LLMs in knowledge-based digital camera recommendation.

Scenario	Example (prompt)	LLM response (excerpt)
Flexibly explaining physical properties	Why does a lower-resolution camera often provide a higher image quality?	In some cases, lower-resolution cameras have larger individual pixels. Larger pixels can capture more light, which can contribute to better image quality, especially in low-light conditions.
Flexible item comparisons	What is the most important difference between the Canon EOS 5d mark IV and the 5d mark II in terms of image quality?	The Mark IV has a wider ISO range and generally better low-light performance compared to the Mark II. The increased resolution does not sacrifice sensitivity, and the Mark IV is capable of producing cleaner images at higher ISO settings.

The examples already show the high flexibility in providing explanations to recommender users.

5.6 Integration of large language models

The application of large language models (LLMs) in knowledge-based recommender systems (and beyond) starts to play a major role in the context of various tasks such as preference elicitation (e.g., in terms of building semantically-enriched user models) and explanations, for example, by identifying the best explanation strategy in specific recommendation contexts (see Table 18—the examples have been generated with ChatGPT 3.5).[11]

5.7 Sustainability aspects of recommender systems

Recommender systems can play a major role in achieving individual sustainability goals. These systems can help, for example, to systematically optimize energy consumption strategies in households, optimize resource reuse in supply chains, and improve personal wellbeing (Murphy et al., 2015; Felfernig et al., 2023b,c). A major issue for future research is to more intensively integrate psychological approaches such as nudging (Thaler and Sunstein, 2021) into recommendation dialogs and to develop new algorithms which help to increase the selection share of sustainable items.

11 openai.com

5.8 Generalizability to other knowledge-based systems

The concepts discussed in this article can be applied to other systems beyond knowledge-based recommender systems. For example, the concepts of personalized constraint-based recommendation and corresponding diagnosis concepts are also applicable in the context of configuration and re-configuration for highly-variant and complex products and services and also in the context of scheduling and re-scheduling (Felfernig et al., 2014).

6 Conclusion

With this article, we provide an overview of major concepts of the knowledge-based recommender systems extending existing overviews with new technological developments. In this context, we provide working examples that help to understand the underlying techniques and algorithms in more detail. Finally, as a result of our literature analysis, we discuss different research directions in knowledge-based recommendation which will help to stimulate further related research.

Author contributions

MU: Conceptualization, Data curation, Formal analysis, Investigation, Project administration, Software, Validation, Writing – original draft, Writing – review & editing. AF: Conceptualization, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing. V-ML: Conceptualization, Validation, Writing – review & editing. TT: Validation, Writing – review & editing. DG: Validation, Writing – review & editing. SL: Validation, Writing – review & editing. TB: Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by TU Graz Open Access Publishing Fund.

Conflict of interest

MU was employed by Siemens Energy AG.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aggarwal, C. C. (2016). *Recommender Systems: The Textbook, 1st Edn.* Springer Nature Switzerland AG.

Almalis, N., Tsihrintzis, G., and Kyritsis, E. (2018). A constraint-based job recommender system integrating fodra. *Int. J. Comp. Intell. Stud.* 7, 103–123. doi: 10.1504/IJCISTUDIES.2018.094894

Atas, M., Felfernig, A., Polat-Erdeniz, S., Popescu, A., Tran, T. N. T., and Uta, M. (2021). Towards psychology-aware preference construction in recommender systems: overview and research issues. *J. Intell. Inf. Syst.* 57, 467–489. doi: 10.1007/s10844-021-00674-5

Bahramian, Z., and Ali Abbaspour, R. (2015). An ontology-based tourism recommender system based on spreading activation model. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 40, 83–90. doi: 10.5194/isprsarchives-XL-1-W5-83-2015

Biere, A., Heule, M., Maaren, H., and Walsh, T. (2021). *Handbook of Satisfiability, 2nd Edn.* Amsterdam: IOS Press.

Bokolo, A. (2021). A case-based reasoning recommender system for sustainable smart city development. *AI Soc.* 36, 159–183. doi: 10.1007/s00146-020-00984-2

Bouraga, S., Jureta, I., Faulkner, S., and Herssens, C. (2014). Knowledge-based recommendation systems: a survey. *Int. J. Intell. Inf. Technol.* 10, 1–19. doi: 10.4018/ijiit.2014040101

Bridge, D., Göker, M. H., McGinty, L., and Smyth, B. (2005). Case-based recommender systems. *Knowl. Eng. Rev.* 20, 315–320. doi: 10.1017/S0269888906000567

Bridge, D. G. (2002). "Towards conversational recommender systems: a dialogue grammar approach," in *ECCBR Workshops* (Aberdeen), 9–22.

Burke, R. (2000). Knowledge-based recommender systems. *Encyclop. Libr. Inf. Syst.* 69, 180–200. Available online at: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=dc133144d431fc3b75c8de27f6bb21da6eb5bc1b

Burke, R. (2002). Hybrid recommender systems: survey and experiments. *User Model. User Adapt. Interact.* 12, 331–370. doi: 10.1023/A:1021240730564

Burke, R., Felfernig, A., and Göker, M. H. (2011). Recommender systems: an overview. *AI Mag.* 32, 13–18. doi: 10.1609/aimag.v32i3.2361

Burke, R., Hammond, K., and Young, B. (1996). Knowledge-based navigation of complex information spaces. *Natl. Conf. Artif. Intell.* 462, 468.

Cena, F., Console, L., and Vernero, F. (2021). Logical foundations of knowledge-based recommender systems: a unifying spectrum of alternatives. *Inf. Sci.* 546, 60–73. doi: 10.1016/j.ins.2020.07.075

Chen, L., and Pu, P. (2012). Critiquing-based recommenders: survey and emerging trends. *User Model User Adapt. Interact.* 22, 125–150. doi: 10.1007/s11257-011-9108-6

Chen, L., Yan, D., and Wang, F. (2017). User perception of sentiment-integrated critiquing in recommender systems. *Int. J. Hum. Comput. Stud.* 121, 4–20. doi: 10.1016/j.ijhcs.2017.09.005

Christakopoulou, K., Radlinski, F., and Hofmann, K. (2016). "Towards conversational recommender systems," in *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16* (New York, NY: ACM), 815–824.

Colombo-Mendoza, L. O., Valencia-García, R., Rodríguez-González, A., Alor-Hernández, G., and Samper-Zapater, J. J. (2015). RecomMetz: a context-aware knowledge-based mobile recommender system for movie showtimes. *Expert Syst. Appl.* 42, 1202–1222. doi: 10.1016/j.eswa.2014.09.016

Cordero, P., Enciso, M., López, D., and Mora, A. (2020). A conversational recommender system for diagnosis using fuzzy rules. *Expert Syst. Appl.* 154:113449. doi: 10.1016/j.eswa.2020.113449

Cöster, R., Gustavsson, A., Olsson, T., and Rudström, Å. (2002). "Enhancing web-based configuration with recommendations and cluster-based help," in *AH'2002*

Workshop on Recommendation and Personalization in eCommerce, 28 May 2002 (Málaga), 1–10.

Dai, S., Shao, N., Zhao, H., Yu, W., Si, Z., Xu, C., et al. (2023). "Uncovering chatgpt?s capabilities in recommender systems," in *17th ACM Conference on Recommender Systems, RecSys '23* (New York, NY: Association for Computing Machinery), 1126–1132.

Daoudi, A., Mechqrane, Y., Bessiere, C., Lazaar, N., and Bouyakhf, E. H. (2016). "Constraint acquisition using recommendation queries," in *International Joint Conferences on Artificial Intelligence* (Palo Alto, CA), 720–726.

Dara, S., Chowdary, C. R., and Kumar, C. (2020). A survey on group recommender systems. *J. Intell. Inf. Syst.* 54, 271–295. doi: 10.1007/s10844-018-0542-3

Dong, M., Zeng, X., Koehl, L., and Zhang, J. (2020). An interactive knowledge-based recommender system for fashion product design in the big data environment. *Inf. Sci.* 540, 469–488. doi: 10.1016/j.ins.2020.05.094

Dragone, P., Teso, S., and Passerini, A. (2018). Constructive preference elicitation. *Front. Robot. AI* 4:71. doi: 10.3389/frobt.2017.00071

Eiter, T., Ianni, G., and Krennwallner, T. (2009). "Answer set programming: a primer," in *Reasoning Web Semantic Technologies for Information Systems: International Summer School*, eds S. Tessaris, E. Franconi, T. Eiter, C. Gutierrez, S. Handschuh, M.-C. Rousset, and R. A. Schmidt (Berlin Heidelberg: Springer), 40–110. doi: 10.1007/978-3-642-03754-2_2

Ekstrand, M., Riedl, J., and Konstan, J. (2011). Collaborative filtering recommender systems. *Found. Trends Human Comp. Interact.* 4. doi: 10.1561/1100000009

Erdeniz, S. P., Felfernig, A., and Atas, M. (2022). Applying matrix factorization to consistency-based direct diagnosis. *Appl. Intell.* 52, 7024–7036. doi: 10.1007/s10489-020-02183-4

Erdeniz, S. P., Felfernig, A., Samer, R., and Atas, M. (2019). "Matrix factorization based heuristics for constraint-based recommenders," in *34th ACM/SIGAPP Symposium on Applied Computing* (New York, NY: ACM), 1655–1662.

Esheiba, L., Elgammal, A., Helal, I. M., and El-Sharkawi, M. E. (2021). A hybrid knowledge-based recommender for product-service systems mass customization. *Information* 12:296. doi: 10.3390/info12080296

Falkner, A., Felfernig, A., and Haag, A. (2011). Recommendation technologies for configurable products. *AI Mag.* 32, 99–108. doi: 10.1609/aimag.v32i3.2369

Fano, A., and Kurth, S. W. (2003). "Personal choice point: Helping users visualize what it means to buy a bmw," in *8th International Conference on Intelligent User Interfaces, IUI '03* (New York, NY: Association for Computing Machinery), 46–52.

Fargier, H., Gimenez, P.-F., and Mengin, J. (2016). "Recommendation for product configuration: an experimental evaluation," in *18th International Configuration Workshop (CWS 2016) within CP 2016: 22nd International Conference on Principles and Practice of Constraint Programming* (Albi: École des Mines d'Albi-Carmaux), 9.

Feely, C., Caulfield, B., Lawlor, A., and Smyth, B. (2020). "Using case-based reasoning to predict marathon performance and recommend tailored training plans," in *Case-Based Reasoning Research and Development: 28th International Conference, ICCBR 2020, Salamanca, Spain, June 8-12, 2020, Proceedings 28* (Springer Nature Switzerland AG), 67–81.

Felfernig, A. (2007). Standardized configuration knowledge representations as technological foundation for mass customization. *IEEE Transact. Eng. Manag.* 54, 41–56. doi: 10.1109/TEM.2006.889066

Felfernig, A., Atas, M., Stettinger, M., Tran, T., and Reiterer, S. (2024a). "Group recommender applications," in *Group Recommender Systems: An Introduction*, eds A. Felfernig, L. Boratto, M. Stettinger, and M. Tkalcic (Springer), 79–94.

Felfernig, A., Boratto, L., Stettinger, M., and Tkalcic, M. (2024b). *Group Recommender Systems: An Introduction*. Springer Publishing Company Inc.

Felfernig, A., and Burke, R. (2008). "Constraint-based recommender systems: technologies and research issues," in *10th International Conference on Electronic Commerce, ICEC '08* (New York, NY: Association for Computing Machinery), 1–10.

Felfernig, A., Friedrich, G., Isak, K., Shchekotykhin, K., Teppan, E., and Jannach, D. (2009a). Automated debugging of recommender user interface descriptions. *Appl. Intell.* 31, 1–14. doi: 10.1007/s10489-007-0105-8

Felfernig, A., Friedrich, G., Jannach, D., and Zanker, M. (2006). An integrated environment for the development of knowledge-based recommender applications. *Int. J. Electron. Commerce* 11, 11–34. doi: 10.2753/JEC1086-4415110201

Felfernig, A., Friedrich, G., Jannach, D., and Zanker, M. (2011). "Developing constraint-based recommenders," in *Recommender Systems Handbook*, eds F. Ricci, L. Rokach, B. Shapira, and P. Kantor (Springer), 187–215.

Felfernig, A., Friedrich, G., Schubert, M., Mandl, M., Mairitsch, M., and Teppan, E. (2009b). "Plausible repairs for inconsistent requirements," in *21st International Joint Conference on Artificial Intelligence, IJCAI'09* (San Francisco, CA. Morgan Kaufmann Publishers Inc.), 791–796.

Felfernig, A., Hotz, L., Bagley, C., and Tiihonen, J. (2014). *Knowledge-based Configuration: From Research to Business Cases*. San Francisco, CA: Morgan Kaufmann Publishers Inc.

Felfernig, A., Isak, K., Szabo, K., and Zachar, P. (2007). "The VITA financial services sales support environment," in *19th National Conference on Innovative Applications of Artificial Intelligence - Volume 2, IAAI'07* (AAAI Press), 1692–1699.

Felfernig, A., and Kiener, A. (2005). "Knowledge-based interactive selling of financial services with FSAdvisor," in *17th Conference on Innovative Applications of Artificial Intelligence - Volume 3, IAAI'05* (AAAI Press), 1475–1482.

Felfernig, A., Le, V., and Lubos, S. (2023a). "Conjunctive query based constraint solving for feature model configuration," in *12th Conference on Information Technology and Its Applications (CITA 2023)*, *volume 734 of Lecture Notes in Networks and Systems* (Springer), 357–367.

Felfernig, A., Le, V.-M., Popescu, A., Uta, M., Tran, T. N. T., and Atas, M. (2021). "An overview of recommender systems and machine learning in feature modeling and configuration," in *15th International Working Conference on Variability Modelling of Software-Intensive Systems, VaMoS '21* (ACM), 1–8.

Felfernig, A., Reiterer, S., Stettinger, M., Reinfrank, F., Jeran, M., and Ninaus, G. (2013a). "Recommender systems for configuration knowledge engineering," in *15th International Configuration Workshop, ConfWS '13*, 51–54.

Felfernig, A., Reiterer, S., Stettinger, M., and Tiihonen, J. (2015). "Towards understanding cognitive aspects of configuration knowledge formalization," in *9th International Workshop on Variability Modelling of Software-Intensive Systems, VaMoS '15* (New York, NY: Association for Computing Machinery), 117–123.

Felfernig, A., Schippel, S., Leitner, G., Reinfrank, F., Isak, K., Mandl, M., et al. (2013b). Automated repair of scoring rules in constraint-based recommender systems. *AI Commun.* 26, 15–27. doi: 10.3233/AIC-120543

Felfernig, A., Schubert, M., and Reiterer, S. (2013c). "Personalized diagnosis for over-constrained problems," in *23rd International Joint Conf. on Artificial Intelligence, IJCAI '13* (AAAI), 1990–1996.

Felfernig, A., Schubert, M., and Zehentner, C. (2012). An efficient diagnosis algorithm for inconsistent constraint sets. *AI EDAM* 26, 53–62. doi: 10.1017/S0890060411000011

Felfernig, A., Stettinger, M., Atas, M., Samer, R., Nerlich, J., Scholz, S., et al. (2018a). "Towards utility-based prioritization of requirements in open source environments," in *2018 IEEE 26th International Requirements Engineering Conference (RE)* (IEEE), 406–411.

Felfernig, A., Walter, R., Galindo, J., Benavides, D., Atas, M., Polat-Erdeniz, S., et al. (2018b). Anytime diagnosis for reconfiguration. *J. Intell. Inf. Syst.* 51, 161–182. doi: 10.1007/s10844-017-0492-1

Felfernig, A., Wundara, M., Tran, T., Le, V.-M., Lubos, S., and Polat-Erdeniz, S. (2023b). Sports recommender systems: overview and research issues. *arXiv*. doi: 10.21203/rs.3.rs-3710874/v1

Felfernig, A., Wundara, M., Tran, T. N. T., Polat-Erdeniz, S., Lubos, S., El Mansi, M., et al. (2023c). Recommender systems for sustainability: overview and research issues. *Front. Big Data* 6:1284511. doi: 10.3389/fdata.2023.1284511

Fesenmaier, D., Ricci, F., Schaumlechner, E., Wöber, K., and Zanella, C. (2003). "Dietorecs: travel advisory for multiple decision styles," in *Information and Communication Technologies in Tourism 2003*, eds A. Frew, M. Hitz, and P. O?Connors (New York, NY: Springer), 232–241.

Friedrich, G., and Zanker, M. (2011). A taxonomy for generating explanations in recommender systems. *AI Mag.* 32, 90–98. doi: 10.1609/aimag.v32i3.2365

Gao, C., Lei, W., He, X., de Rijke, M., and Chua, T.-S. (2021). Advances and challenges in conversational recommender systems: a survey. *AI Open* 2, 100–126. doi: 10.1016/j.aiopen.2021.06.002

Goeker, M., and Thompson, C. (2000). "Personalized conversational case-based recommendation," in *Advances in Case-Based Reasoning, Volume 1898 of Lecture Notes in Computer Science*, eds E. Blanzieri, and L. Portinale (Berlin; Heidelberg: Springer), 709–714.

Grasch, P., Felfernig, A., and Reinfrank, F. (2013). "ReComment: towards critiquing-based recommendation with speech interaction," in *7th ACM Conference on Recommender Systems, RecSys '13* (New York, N: ACM), 157–164.

Güell, M., Salamó, M., Contreras, D., and Boratto, L. (2020). Integrating a cognitive assistant within a critique-based recommender system. *Cogn. Syst. Res.* 64:1–14. doi: 10.1016/j.cogsys.2020.07.003

Hernandez-Nieves, E., Hernández, G., Gil-Gonzalez, A. B., Rodríguez-González, S., and Corchado, J. M. (2021). Cebra: a case-based reasoning application to recommend banking products. *Eng. Appl. Artif. Intell.* 104:104327. doi: 10.1016/j.engappai.2021.104327

Iaquinta, L., deGemmis, M., Lops, P., Semeraro, G., Filannino, M., and Molino, P. (2008). "Introducing serendipity in a content-based recommender system," in *8th International Conference on Hybrid Intelligent Systems, HIS '08* (Barcelona: IEEE Computer Society), 168–173.

Jannach, D. (2006). "Techniques for fast query relaxation in content-based recommender systems," in *29th Annual German Conference on Artificial Intelligence* (Berlin, Heidelberg: Springer), 49–63.

Jannach, D., and Kreutler, G. (2007). Rapid development of knowledge-based conversational recommender applications with advisor suite. *J. Web Eng.* 6, 165–192. Available online at: https://www.researchgate.net/profile/Dietmar-Jannach/publication/220538223_Rapid_Development_of_Knowledge-Based_Conversational_Recommender_Applications_with_Advisor_Suite/links/5e30386a299bf1cdb9f69472/Rapid-Development-of-Knowledge-Based-Conversational-Recommender-Applications-with-Advisor-Suite.pdf?_sg%5B0%5D=started_experiment_milestone&origin=journalDetail

Jannach, D., Manzoor, A., Cai, W., and Chen, L. (2021). A survey on conversational recommender systems. *ACM Comp. Surv.* 54, 1–36. doi: 10.1145/3453154

Junker, U. (2004). "QuickXPlain: Preferred explanations and relaxations for over-constrained problems," in *19th National Conference on Artifical Intelligence, AAAI'04* (AAAI Press), 167–172.

Khan, A., and Hoffmann, A. (2003). Building a case-based diet recommendation system without a knowledge engineer. *Artif. Intell. Med.* 27, 155–179. doi: 10.1016/S0933-3657(02)00113-6

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., and Newell, C. (2012). Explaining the user experience of recommender systems. *User Model. User Adapt. Interact.* 22, 441–504. doi: 10.1007/s11257-011-9118-4

Kolodner, J. (2014). *Case-Based Reasoning*. San Mateo, CA: Morgan Kaufmann.

Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artif. Intell. Rev.* 6, 3–34. doi: 10.1007/BF00155578

Law, E., and von Ahn, L. (2011). *Human Computation, 1st Edn*. Kentfield, CA: Morgan & Claypool Publishers.

Le, V., Felfernig, A., Uta, M., Benavides, D., Galindo, J., and Tran, T. (2023). "FASTDIAGP: an algorithm for parallelized direct diagnosis," in *AAAI Conference on Artificial Intelligence*, 6442–6449.

Le, V.-M., Felfernig, A., Uta, M., Benavides, D., Galindo, J., and Tran, T. N. T. (2021). "DirectDebug: automated testing and debugging of feature models," in *43rd International Conference on Software Engineering: New Ideas and Emerging Results, ICSE-NIER '21* (Madrid: IEEE Press), 81–85.

Le, V.-M., Tran, T. N. T., and Felfernig, A. (2022). "Consistency-based integration of multi-stakeholder recommender systems with feature model configuration," in *26th ACM International Systems and Software Product Line Conference-Volume B*, 178–182.

Lee, H. J., and Kim, H. S. (2015). "eHealth recommendation service system using ontology and case-based reasoning," in *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)* (Chengdu: IEEE), 1108–1113.

Lee, T., Chun, J., Shim, J., and Lee, S. (2006). An ontology-based product recommender system for B2B marketplaces. *Int. J. Electron. Commerce* 11, 125–155. doi: 10.2753/JEC1086-4415110206

Lika, B., Kolomvatsos, K., and Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Syst. Appl.* 41, 2065–2073. doi: 10.1016/j.eswa.2013.09.005

Lorenzi, F., and Ricci, F. (2005). "Case-based recommender systems: a unifying view," in *Intelligent Techniques for Web Personalization, Volume 3169 of Lecture Notes in Computer Science*, eds B. Mobasher and S. Anand (Berlin; Heidelberg: Springer), 89–113.

Lubos, S., Le, V.-M., Felfernig, A., and Tran, T. N. T. (2023). "Analysis operations for constraint-based recommender systems," in *17th ACM Conference on Recommender Systems, RecSys '23* (New York, NY: ACM), 709–714.

Maccatrozzo, V., Terstall, M., Aroyo, L., and Schreiber, G. (2017). "Sirup: serendipity in recommendations via user perceptions," in *22nd International Conference on Intelligent User Interfaces, IUI '17* (New York, NY: ACM), 35–44.

Mandl, M., and Felfernig, A. (2012). "Improving the performance of unit critiquing," in *20th International Conference on User Modeling, Adaptation, and Personalization, UMAP'12* (Berlin, Heidelberg: Springer-Verlag), 176–187.

Masthoff, J. (2015). "Group recommender systems: aggregation, satisfaction and group attributes," in *Recommender Systems Handbook, Chapter 22* (New York, NY: Springer Science+Business Media), 743–776.

McCarthy, K., McGinty, L., Smyth, B., and Salam,ó, M. (2006). "The needs of the many: a case-based group recommender system," in *ECCBR 2006, Volume 4106 of Lecture Notes in Computer Science*, eds T. Roth-Berghofer, M. Göker, and H. Güvenir (Berlin; Heidelberg: Springer), 196–210.

McCarthy, K., Reilly, J., McGinty, L., and Smyth, B. (2005). "Experiments in dynamic critiquing," in *10th International Conference on Intelligent User Interfaces, IUI '05* (New York, NY: Association for Computing Machinery), 175–182.

McCarthy, K., Salem, Y., and Smyth, B. (2010). "Experience-based critiquing: reusing critiquing experiences to improve conversational recommendation," in *ICCBR'10* (Springer), 480–494.

McGuinness, D. (2007). "Configuration," in *The Description Logic Handbook: Theory, Implementation and Applications, 2 Edn* (Cambridge: Cambridge University Press), 417–435.

McSherry, D. (2003). "Similarity and compromise," in *Case-Based Reasoning Research and Development*, eds K. D. Ashley, and D. G. Bridge (Berlin, Heidelberg: Springer Berlin Heidelberg), 291–305.

Mirzadeh, N., Ricci, F., and Bansal, M. (2005). "Feature selection methods for conversational recommender systems," in *2005 IEEE International Conference on E-Technology, e-Commerce and e-Service (EEE'05) on e-Technology, e-Commerce and e-Service, EEE '05* (Hong Kong: IEEE Computer Society), 772–777.

Monroe, D. (2022). Neurosymbolic AI. *Commun. ACM* 65, 11–13. doi: 10.1145/3554918

Murphy, S. Ó., Manzano, Ó., and Brown, K. N. (2015). "Design and evaluation of a constraint-based energy saving and scheduling recommender system," in *International Conference on Principles and Practice of Constraint Programming* (Springer International Publishing Switzerland), 687–703.

Murti, Y., A., Baizal, Z., and Adiwijaya, K. (2016). Compound critiquing for conversational recommender system based on functional requirement. *Adv. Sci. Lett.* 22, 1892–1896. doi: 10.1166/asl.2016.7770

Musto, C., Semeraro, G., Lops, P., De Gemmis, M., and Lekkas, G. (2015). Personalized finance advisory through case-based recommender systems and diversification strategies. *Decis. Support Syst.* 77, 100–111. doi: 10.1016/j.dss.2015.06.001

Pazzani, M. J., and Billsus, D. (2007). "Content-based recommendation systems," in *The Adaptive Web: Methods and Strategies of Web Personalization* (Berlin; Heidelberg: Springer), 325–341.

Pereira, J., Matuszyk, P., Krieter, S., Spiliopoulou, M., and Saake, G. (2016). A feature-based personalized recommender system for product-line configuration. *SIGPLAN Not.* 52:2993249. doi: 10.1145/2993236.2993249

Pessemier, T. D., Dhondt, J., and Martens, L. (2017). Hybrid group recommendations for a travel service. *Multimed. Tools Appl.* 76, 2787–2811. doi: 10.1007/s11042-016-3265-x

Popescu, A., Polat-Erdeniz, S., Felfernig, A., Uta, M., Atas, M., Le, V.-M., et al. (2022). An overview of machine learning techniques in constraint solving. *J. Intell. Inf. Syst.* 58, 91–118. doi: 10.1007/s10844-021-00666-5

Pramod, D., and Bafna, P. (2022). Conversational recommender systems techniques, tools, acceptance, and adoption: a state of the art review. *Expert Syst. Appl.* 203:117539. doi: 10.1016/j.eswa.2022.117539

Rafter, R., and Smyth, B. (2005). Conversational collaborative recommendation—An experimental analysis. *Artif. Intell. Rev.* 24, 301–318. doi: 10.1007/s10462-005-9004-8

Reilly, J., McCarthy, K., McGinty, L., and Smyth, B. (2004). "Dynamic critiquing," in *Advances in Case-Based Reasoning: 7th European Conference, ECCBR 2004, Madrid, Spain, August 30-September 2, 2004. Proceedings 7* (Berlin; Heidelberg: Springer), 763–777.

Reiter, R. (1987). A theory of diagnosis from first principles. *Artif. Intell.* 32, 57–95. doi: 10.1016/0004-3702(87)90062-2

Rossi, F., van Beek, P., and Walsh, T. (2006). *Handbook of Constraint Programming*. Amsterdam: Elsevier Science.

Sha, X., Sun, Z., and Zhang, J. (2021). Hierarchical attentive knowledge graph embedding for personalized recommendation. *Electron. Commer. Res. Appl.* 48:101071. doi: 10.1016/j.elerap.2021.101071

Smyth, B., McGinty, L., Reilly, J., and McCarthy, K. (2004). "Compound critiques for conversational recommender systems," in *IEEE/WIC/ACM International Conference on Web Intelligence (WI'04)* (Beijing: IEEE), 145–151.

Sun, R., Cao, X., Zhao, Y., Wan, J., Zhou, K., Zhang, F., et al. (2020). "Multi-modal knowledge graphs for recommender systems," in *29th ACM International Conference on Information & Knowledge Management*, 1405–1414.

Tazl, O. A., Perko, A., and Wotawa, F. (2019). "Conversational recommendations using model-based reasoning," in *CEUR Workshop Proceedings, Vol. 2467* (Otzenhausen: RWTH Aachen), 13–19.

Temple, P., Acher, M., Jézéquel, J.-M., and Barais, O. (2017). Learning contextual-variability models. *IEEE Softw.* 34, 64–70. doi: 10.1109/MS.2017.4121211

Teppan, E., and Zanker, M. (2020). "Exploiting answer set programming for building explainable recommendations," in *ISMIS 2020, volume 12117 of Lecture Notes in Computer Science* (Springer), 395–404.

Thaler, R., and Sunstein, C. (2021). *Nudge: The Final Edition*. New Haven, CT; London: Yale University Press.

Tintarev, N., and Masthoff, J. (2012). Evaluating the effectiveness of explanations for recommender systems. *User Model. User Adapt. Interact.* 22, 399–439. doi: 10.1007/s11257-011-9117-5

Towle, B., and Quinn, C. (2000). "Knowledge based recommender systems using explicit user models," in *AAAI Technical Report WS-00-04* (AAAI Press), 74–77.

Tran, T., Felfernig, A., and Le, V. (2023). An overview of consensus models for group decision-making and group recommender systems. *User Model. User Adapt. Interact*. doi: 10.1007/s11257-023-09380-z

Ulz, T., Schwarz, M., Felfernig, A., Haas, S., Shehadeh, A., Reiterer, S., et al. (2017). Human computation for constraint-based recommenders. *J. Intell. Inf. Syst*. 49, 37–57. doi: 10.1007/s10844-016-0433-4

Uta, M., Felfernig, A., Helic, D., and Le, V.-M. (2022). "Accuracy- and consistency-aware recommendation of configuratiosn," in *26th ACM International Systems and Software Product Line Conference - Volume A, SPLC '22* (New York, NY: Association for Computing Machinery), 79–84.

Uta, M., Felfernig, A., Le, V., Popescu, A., Tran, T., and Helic, D. (2021). "Evaluating recommender systems in feature model configuration," in *25th ACM International Systems and Software Product Line Conference - Volume A, SPLC '21* (New York, NY: Association for Computing Machinery), 58–63.

Walter, R., Felfernig, A., and Küchlin, W. (2017). Constraint-based and SAT-based diagnosis of automotive configuration problems. *J. Intell. Inf. Syst*. 49, 87–118. doi: 10.1007/s10844-016-0422-7

Wang, H., Zhang, F., Zhang, M., Leskovec, J., Zhao, M., Li, W., et al. (2019a). "Knowledge-aware graph neural networks with label smoothness regularization for recommender systems," in *25th ACM SIGKDD International Conference on Knowledge Discovery* & *Data Mining* (New York, NY: ACM), 968–977.

Wang, H., Zhao, M., Xie, X., Li, W., and Guo, M. (2019b). "Knowledge graph convolutional networks for recommender systems," in *The World Wide Web Conference* (New York, NY: ACM), 3307–3313.

Weckesser, M., Kluge, R., Pfannemüller, M., Matthé, M., Schürr, A., and Becker, C. (2018). "Optimal reconfiguration of dynamic software product lines based on performance-influence models," in *22nd International Systems and Software Product Line Conference, Vol. 1* (New York, NY: ACM), 98–109.

Wobcke, W., Krzywicki, A., Kim, Y. S., Cai, X., Bain, M., Compton, P., et al. (2015). A deployed people-to-people recommender system in online dating. *AI Mag*. 36, 5–18. doi: 10.1609/aimag.v36i3.2599

Wohlin, C. (2014). "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *18th Intl. Conference on Evaluation and Assessment in Software Engineering* (New York, NY: ACM), 1–10.

Wu, G., Luo, K., Sanner, S., and Soh, H. (2019). "Deep language-based critiquing for recommender systems," in *13th ACM Conference on Recommender Systems, RecSys '19* (New York, NY: Association for Computing Machinery), 137–145.

Xie, H., Wang, D. D., Rao, Y., Wong, T.-L., Raymond, L. Y., Chen, L., et al. (2018). Incorporating user experience into critiquing-based recommender systems: a collaborative approach based on compound critiquing. *Int. J. Mach. Learn. Cybernet*. 9, 837–852. doi: 10.1007/s13042-016-0611-2

Zanker, M., Bricman, M., Gordea, S., Jannach, D., and Jessenitschnig, M. (2006). "Persuasive online-selling in quality and taste domains," in *7th International Conference on E-Commerce and Web Technologies, EC-Web'06* (Berlin, Heidelberg: Springer-Verlag), 51–60.

Zhang, J., Jones, N., and Pu, P. (2008). "A visual interface for critiquing-based recommender systems," in *9th ACM Conference on Electronic Commerce* (New York, NY: ACM), 230–239.

Zhou, K., Zhao, W. X., Bian, S., Zhou, Y., Wen, J.-R., and Yu, J. (2020). "Improving conversational recommender systems via knowledge graph based semantic fusion," in *26th ACM SIGKDD International Conference on Knowledge Discovery* & *Data Mining* 1006–1014.

Zhu, Q., Zhou, X., Wu, J., Tan, J., and Guo, L. (2020). "A knowledge-aware attentional reasoning network for recommendation," in *AAAI Conference on Artificial Intelligence* (Palo Alto, CA: AAAI), 6999–7006.

Ziarani, R., and Ravanmehr, R. (2021). Serendipity in recommender systems: a systematic literature review. *J. Comput. Sci. Technol.* 36, 375–396. doi: 10.1007/s11390-020-0135-9

Zou, J., Chen, Y., and Kanoulas, E. (2020). "Towards question-based recommender systems," in *43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '20* (New York, NY: ACM), 881–890.

Frontiers in Big Data

Explores the potential for big data to address global challenges

This innovative journal focuses on the power of big data - its role in machine learning, AI, and data mining, and its practical application from cybersecurity to climate science and public health.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact